

How to Design a Dashboard



How to Design a Dashboard

Written by: [Matt David](#)

Reviewed by: [Tim Miller](#)

Table of Contents

Introduction

- [Introduction](#)
- [What is a Dashboard?](#)
- [What Makes a Great Dashboard \(ACES\)](#)
- [Dashboard Design Process](#)

Define

- [Identifying Key Roles](#)
- [Determine the Metrics to Monitor](#)

Prototype

- [Find the Best Visualizations for Your Metrics](#)
- [Arranging Your Charts as a Dashboard](#)
- [Dashboard Prototyping and Feedback](#)

Build

- [Finding the Data That Builds Metrics](#)
- [Build the Metrics](#)

Deploy

- [Sharing the Dashboard – Distribution Strategies](#)
- [Scaling Dashboards](#)
- [Making Sure Your Dashboard Always Gets Better](#)

Conclusion

- [Conclusion](#)
- [Design a Dashboard Example](#)

Introduction

Introduction

“The greatest value of a picture is when it forces us to notice what we never expected to see.” —John Tukey, Mathematician

Without looking at data how do people make decisions?

They base it on their past experience and their understanding of the scenario. This is no longer enough as more organizations become data driven. While we cannot change people’s past experience we can surface data to help change their understanding of the scenario.

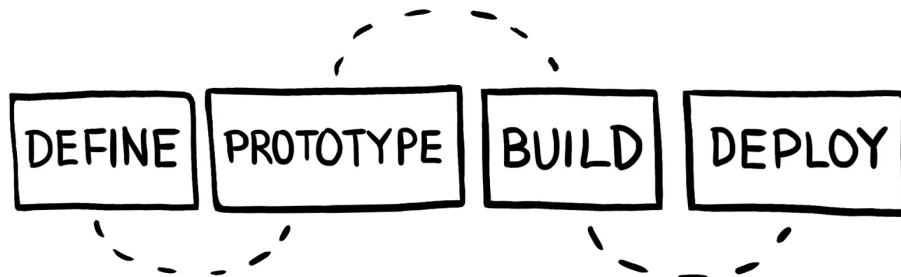
We live in exciting times for data driven decision making:

- We are able to get data from all parts of our business
- We can store tons of data cheaply
- Many tools exist to easily extract and visualize that data

We want more people within our organizations to better understand what is going on so they can make better decisions. We can do this by exposing people to information dashboards. Dashboards are the link between the data people (people like you) and the business people. This book shows how design thinking and design processes can be used to create highly impactful dashboards to help business people make data driven decisions in your organization.

Business Intelligence tools have made it easy to create visualizations and dashboards. It is tempting to start building multiple dashboards right away without fully investing in defining the problems, stakeholders, and metrics or prototyping ideal designs. Spending time in the Define and Prototyping stages will help any dashboard designer produce dashboards that get used more by their audience.

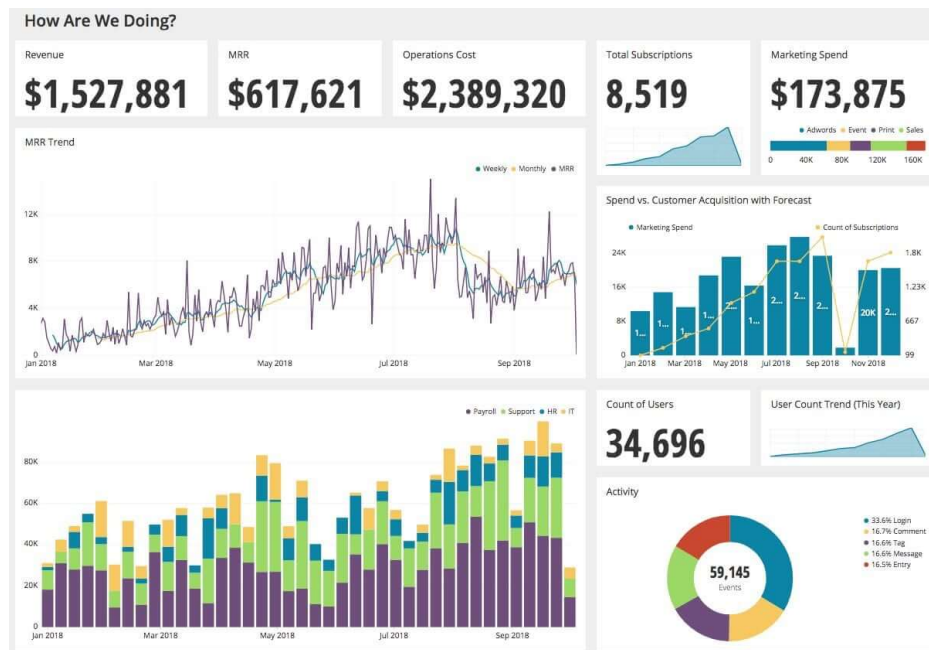
This book will quickly introduce you to what dashboards are, what makes them useful, and an overview of the dashboard design process. Then it will spend the bulk of the book going through the process itself:



This book will provide resources and examples to aid you at every step of the process. Use this book to improve your own dashboard skills and use it as a reference for new analysts that join your organization.

What is a Dashboard?

Anytime you want to display information to help people make decisions you are in the process of creating a dashboard. Dashboards allow us to arrange multiple data visualizations to give people enough context to consistently make great decisions.



For example this is a dashboard tracking the top metrics for a SaaS company. It provides at a glance information related to revenue, operating costs, total users, and other relevant data that can be evaluated.

This dashboard can help the CEO or anyone in the company figure out *what is going on at a high level*, and help him decide where to take action.

“My operations costs are higher than my revenue, I need to reach out to my COO to get informed about why we are spending so much.”

Decision assisted, good job dashboard

Dashboards are built to trigger insights that help you take action, *in the case shown above* the data indicates that an email needs to be sent to someone in operations. This dashboard is composed of various data visualizations to provide the viewer context to support insights and make decisions. A dashboard is dynamic because as the underlying data changes, the dashboard is automatically updated so that time sensitive insights and decisions can be made.

This book will explore best practices to create useful dashboards such as this one to help individuals make data-driven decisions.

History

The term “[dashboard](#)” originates from a board that was built into carriages to block the dirt that the horse dash-ed up. When carriages became automobiles, the dashboard remained relevant to block dirt dash-ed up by the front wheel. When the design of automobiles shifted to putting the engine in the front the dashboard’s purpose grew to protecting the driver from the heat and oil. The dashboard also became a convenient location to put gauges to monitor the engine’s performance and other data points about the car such as fuel levels. The term dashboard jumps to describing the modern day business tool this book is all about due to Microsoft. Microsoft is given credit for promoting the term as part of their Digital Nervous System concept in the 90s.

The purpose of the modern day business dashboard has its origins in research that started in the 1970s to use computers to help people make better decisions. Originally known as Decisions Support Systems they became initially commercialized as Executive Information

Systems. Now these tools are ubiquitous in business to track performance and help people make informed decisions.

Dashboard Value Proposition

Let's begin with that very common dashboard that billions of people use every day, the dashboard in a car. Think about the type of data that is displayed. This dashboard presents how fast the car is going, the RPM, the oil temperature, and how much gas is left. Not to mention warning symbols, information about gears, and whether or not your lights are on. It provides all things you need to monitor in order to do your job, in this case driving your car.

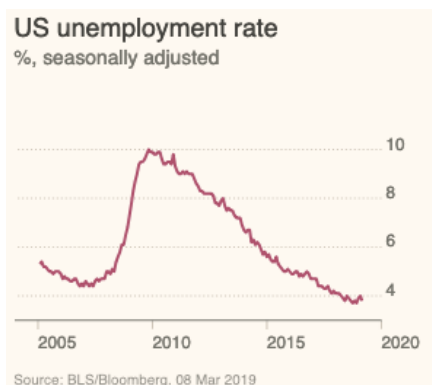


The data displayed to the driver helps them make important decisions about speed, engine health, and if they need to go to a gas station. Simple dashboards such as this allow people to make informed decisions. However, let's imagine a dashboard in a car that only showed one of these data visualizations.



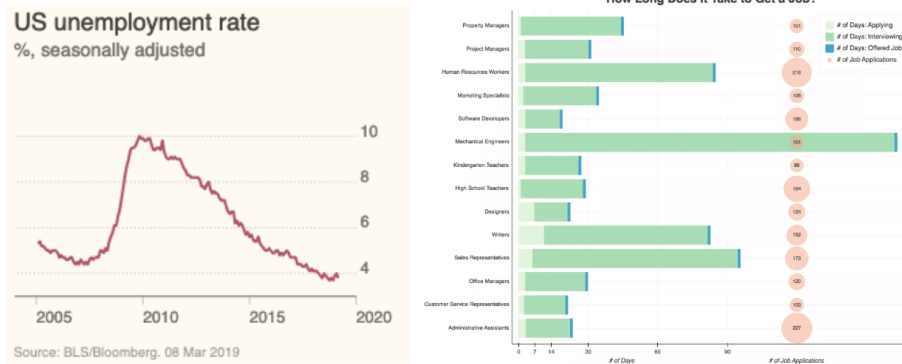
We might know how fast we are going but we could be redlining our engine or we could run out of gas at any minute. Seeing a single chart to make all of your decisions from is dangerous. Unfortunately, this is often how people encounter data, in a single chart. The data is isolated, lacking context and other important information required to make a good decision.

A news report might show the job market of the United States in a single chart:



At first glance this looks great, the unemployment rate is around 4%. This might mean it is easy to find a job. This graph could be useful if I'm considering whether I should look for a better job. One visualization alone however should not give me much confidence. There could be critically important data that is missing by considering only one visualization.

If this was part of a dashboard that included more visualizations about the job market, it would be easier to make a better decision:



While this is not a full dashboard we can see the value in having a second visualization. Placing two graphs next to each other, helps the viewer to see that even though unemployment is low, there are some jobs that take a lot of time and effort to get hired. If I want to get a new job quickly, I should consider software development (not a shocker). Dashboards are composed of multiple visualizations in a single window so all the relevant information is available at once and can be simultaneously evaluated.

Why are dashboards more important than ever?

Software is eating the world. It does this by turning everything it can into data, processing that data, and distributing that data with more efficiency than ever before. Once software digitizes something into data, that data can be leveraged by individuals to make higher quality decisions.

Google helps us make decisions about where to purchase delicious Dim Sum. Netflix helps us make decisions about the next show to binge watch. Tinder helps us make decisions about who to date (questionable :)). These decision-making tools exist inside of organizations as well to help them operate effectively and efficiently, however instead of iPhone and Android applications, they come in the form of dashboards built in business intelligence tools.

OK GOOGLE,

Google

TELL ME HOW TO INCREASE PROFITS

People have become accustomed to relying on data to make decisions in their personal life and want the same thing in their companies. In the past, it was challenging to track and store data. Today storage is incredibly cheap, capturing data is ubiquitous, and connecting all these different sources into a database is relatively simple. This has created an incredible opportunity to make more informed decisions in your company based on relevant and timely data from several sources in context.

How do dashboards get used

Ad Hoc Analysis

When a question comes into your head about how your business is performing you want to get the data and visualize it. This rarely involves any amount of planning or design as you begin writing your query and choosing simple visualizations to look at the data. You may choose to explore the question in a few different ways which means that multiple (unoptimized) visualizations are brought into the dashboard. This dashboard typically has an audience of one (you) or when shared with others is accompanied with a write up or verbal explanation for how to interpret it.

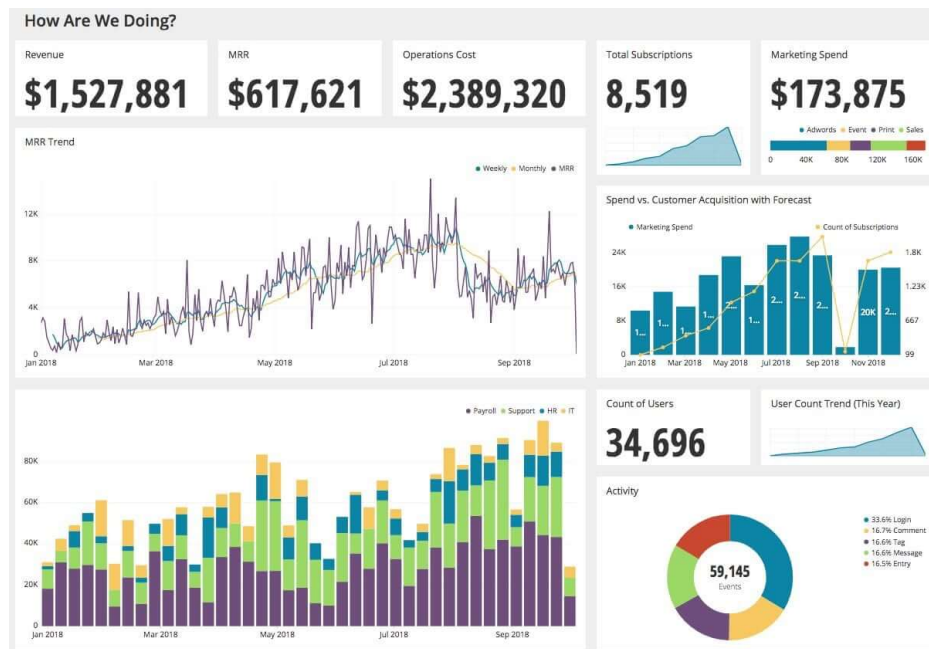
Reports

Often times there will be a large project or large decision that needs historical data laid out so that it can be evaluated. Since it is a static view of the data it typically only provides value for the specific scenario. If this data needs to be evaluated repeatedly then it moves into a decision support dashboard which requires more design and iteration.

Ongoing Decision Support

The strongest use case for dashboards are one that are created to support ongoing decision making. In this way they are built with the same ethos of modern software development using design thinking and iteration. The data powering the visualizations is regularly updated and the dashboard is designed around the audience who will be using it, often times being updated as use cases shift. Ongoing decision support dashboards will be the focus throughout the rest of the book.

Let's revisit the original dashboard which is a decision support dashboard:



This dashboard displays many types of data from different sources:

- Revenue - Payment provider such as Stripe
- Cost - Accounting software such as Quickbooks
- User Count and Activity - Your production database using tools such as Segment

This dashboard includes some predictive measures and groups data in intelligent ways.

Decision makers seek data that provides this level of insight into what is happening in their organizations. They want to make informed decisions based on data. Organizations have more data available than ever before and dashboards such as this one, allow people to leverage all of the data in context for effective decision making. This book will guide you in constructing useful dashboards for any organization.

Summary

- Dashboards are a dynamic display of information to support high quality decision making
- People leverage data all the time to make important decisions in their personal lives and in their organizations
- There is more data available than ever before within organizations
- Dashboards help individuals make informed decisions based on multiple sources of data within an organization

What Makes a Great Dashboard (ACES)

An optimal dashboard is Accurate, Clear, Empowering, and Succinct. These key tenets can be remembered with the acronym ACES.

Accurate

A dashboard lives and dies by the trust the viewers have in what they are seeing. If the viewers doubt the accuracy it will not be used to make decisions. People will also be more hesitant to trust any dashboard or to do any querying themselves. A lack of accuracy in one dashboard causes a lack of faith in all the data.

Viewers belief in the Accuracy of the dashboard can be affected in multiple ways:

- Data Quality
- Metric Comprehension
- Visualization Design

Data Quality

Is the data being displayed correct?

The answer to this question should always be yes. If for some reason the answer is no immediately flag the dashboard as needing to be fixed so people do not use incorrect information for their decisions. Viewers will assume any dashboard they come across to be accurate unless properly flagged otherwise. Use bracketed language [BROKEN] or perhaps emoji's to make the status of the dashboard clear.

✖ Example Dashboard [BROKEN] ✖

Is the data being displayed all of the data?

Most of the times it is not, this is because of how data is loaded into the data warehouse. Engineers use batch processing that runs on a schedule to load data from their production database to the data warehouse which is what the dashboard is querying. This can cause people confusion who are dealing with customers or scenarios in real time where they are not seeing the data in the dashboard. If the dashboard is not displaying all of the data due to batch processing, you should note on your dashboard when the data was last updated and it's schedule.

Updated 10 hours ago

Updates every 12 hours

Metric Comprehension

Metrics need to be understood before the viewer can interpret the chart accurately. It is a best practice to include formulas, notes, or definitions for any non-traditional metric directly on the dashboard. Placing it next to the visualization using it allows for the quickest use. Here we can easily see clarification around who is not included in this metric.

Net New Trials

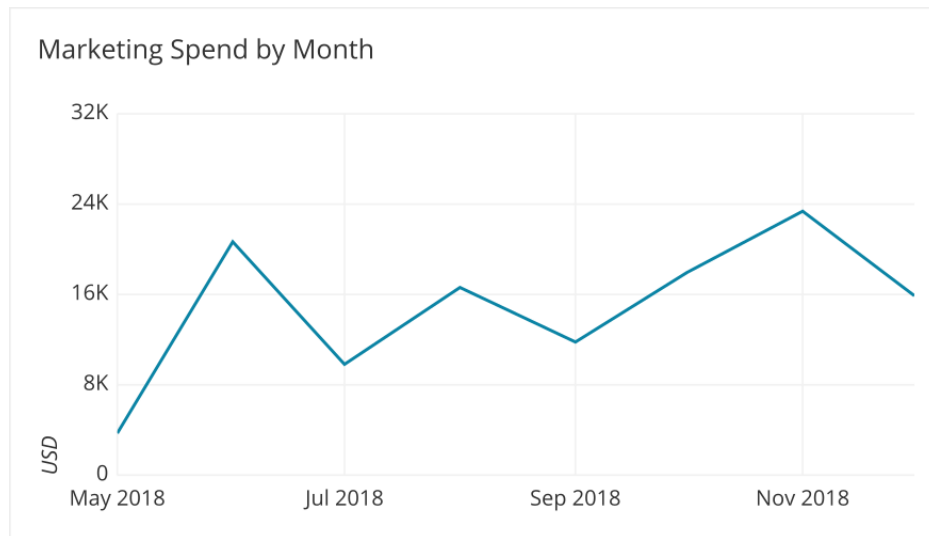
946

Not measuring trials from existing customers or those that have not completed their password.

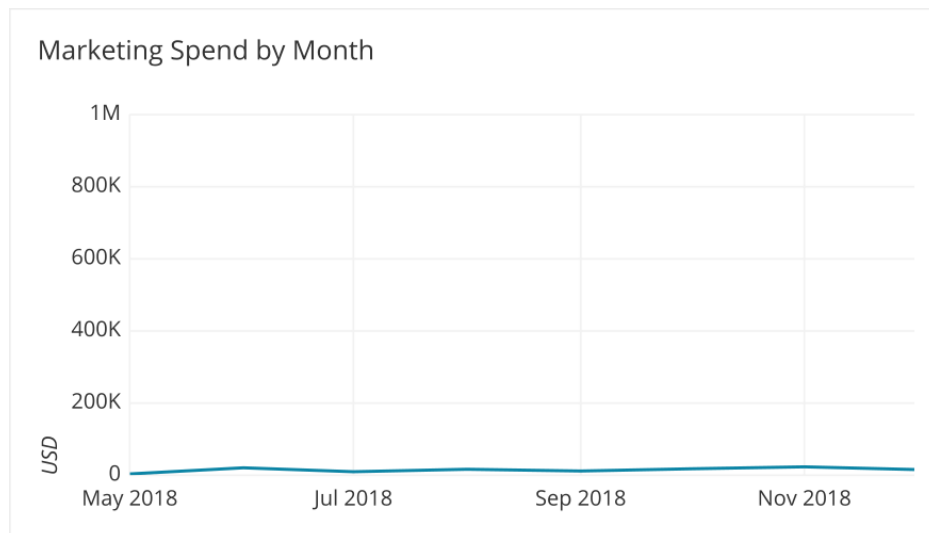
Visualization Design

People are visual creatures and have a lot of biases in interpreting any visualization of data. A common mistake is setting the Y-axis range incorrectly. We want to highlight variation but we do not want to bias interpretation. For line graphs we do not have to start at 0 and we want to capture how the data changes

Good Y-axis Range - Can clearly see the variability

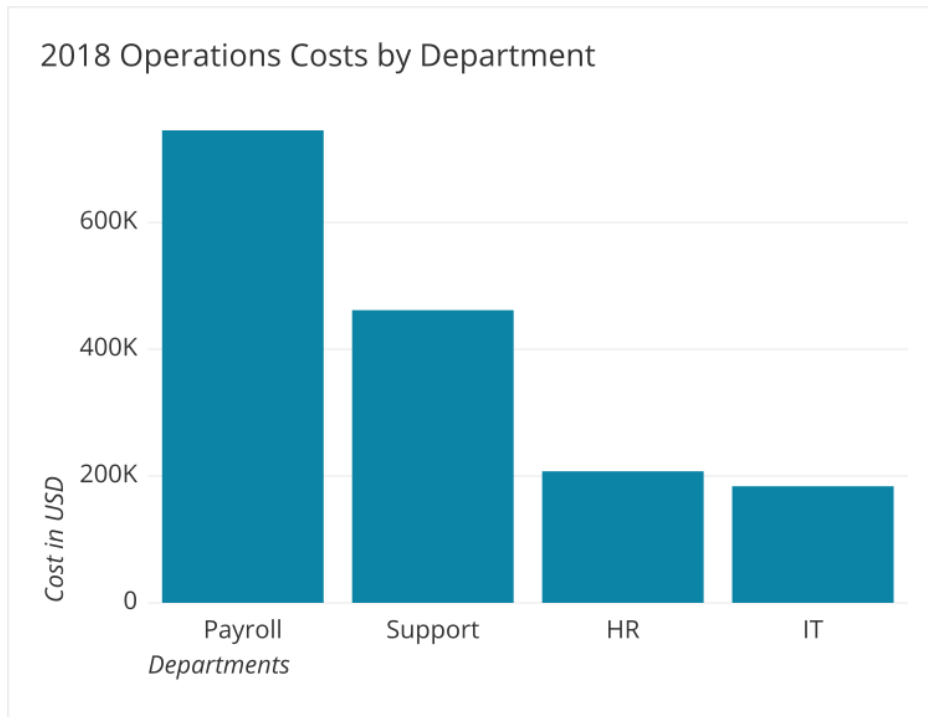


Bad Y-axis Range - Cannot see the variability

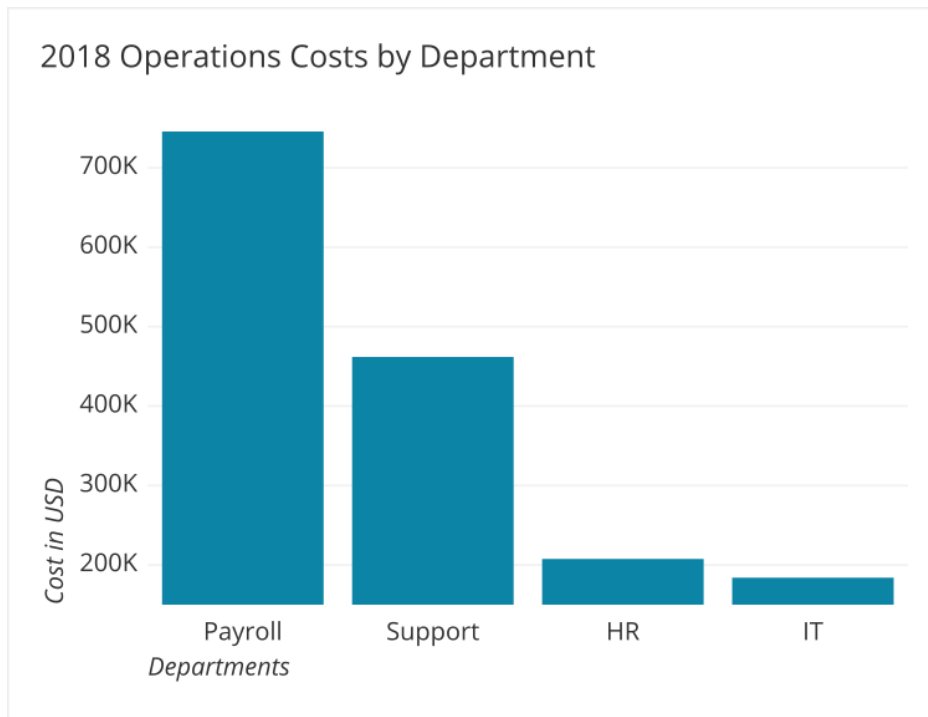


For bar graphs we must start at 0 because if it starts at a different point it prevents us from being able to use the size to judge the difference. When we look at the two examples below in the first we can see that HR is about half of what support is which is correct Support is ~450,000 and HR is ~200,000. However when we look at the second example HR looks like it is 1/5 of what Support is which is not correct.

Good Y-axis Range - Starts at 0



Bad Y-axis Range - Does not start at 0



Clear

To be able to make a decision based off of a dashboard, the data must be displayed clearly. There are several factors that go into making the data clear.

- Fonts
- Colors
- Context
- Layout

Fonts

Fonts for chart titles, axes labels, and details should not be decorative. The goal is legibility. We recommend using a sans serif font such as Arial, Helvetica, or Verdana.

Fancy Chart Title ✖
Legible Chart Title ✔

Consider the font size of any metric that will be displayed on your dashboard. Think about what type of device the audience will be viewing this on as well. All text on a dashboard should not require you to squint to read it.

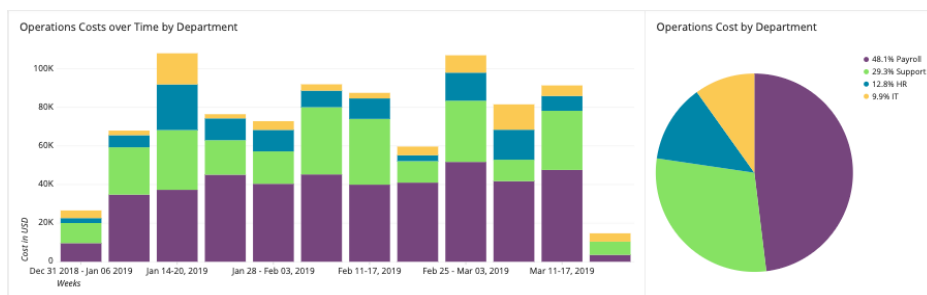
Total Subscriptions
8,519 VS **8,519**

Colors

The color palette used for the dashboard should be easy on the eyes and not overweight one color over another. An easy way to accomplish this is to use more muted tones.

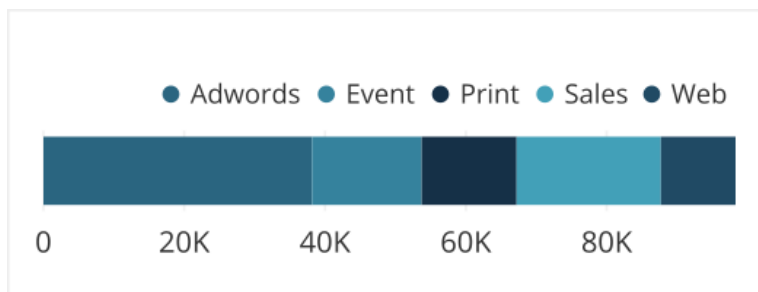


Colors should also be consistent when representing the same metric or type of metric from chart to chart. This makes it easy to relate the data across graphs, tables and charts.

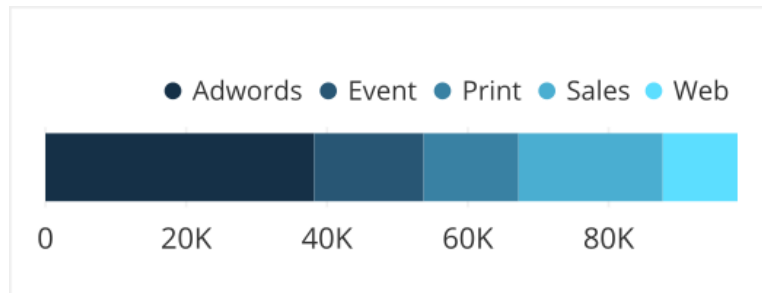


The colors used in the visualizations need to contrast the background enough to be seen clearly. However, too much contrast within a chart can be distracting.

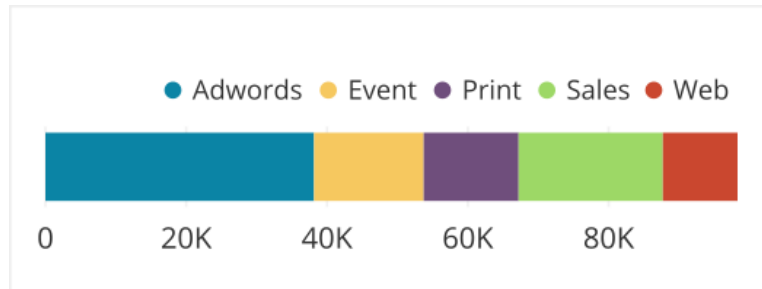
We can examine some different color palette choices below and better alternatives.



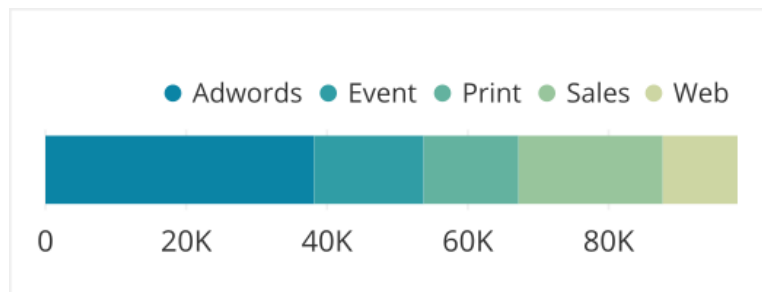
This is difficult to see which part of the bar belongs to which category



This shows the difference between the categories but the last part of the chart is too saturated and grabs your attention so this is not ideal unless you are trying to highlight that section.



This differentiation is clear but some of the colors are very far apart on the color wheel and could become a bit distracting if these colors are used too much on the dashboard.



These colors show the difference between the categories without overweighting one of the categories. Most dashboarding tools will default you to a color scheme that delineates categories clearly. If for some reason they do not, then customize your color selection so that evaluations of the data can be made at a glance.

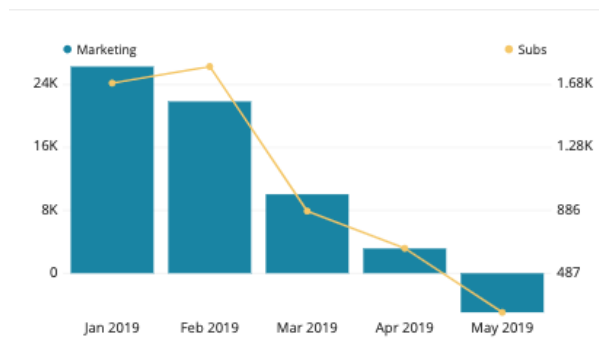
[The Data Color Picker tool](#) is a great resource for picking evenly spaced out colors for any visualization.

Context

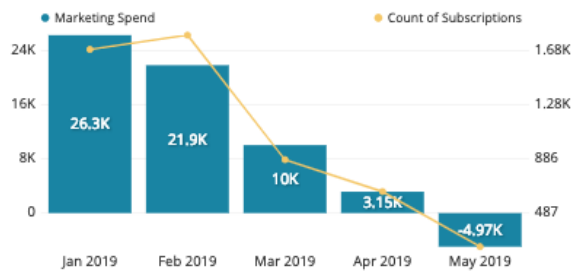
Include information such as:

- A descriptive title
- Categorical labels
- Value labels when it is hard to compare against an axis

These additions make it easier for new viewers to understand what is going on.



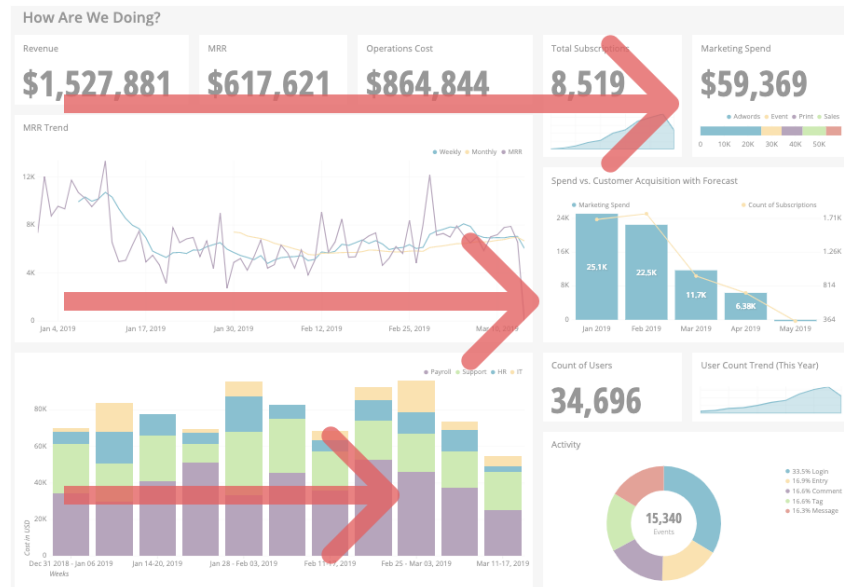
Spend vs. Customer Acquisition with Forecast



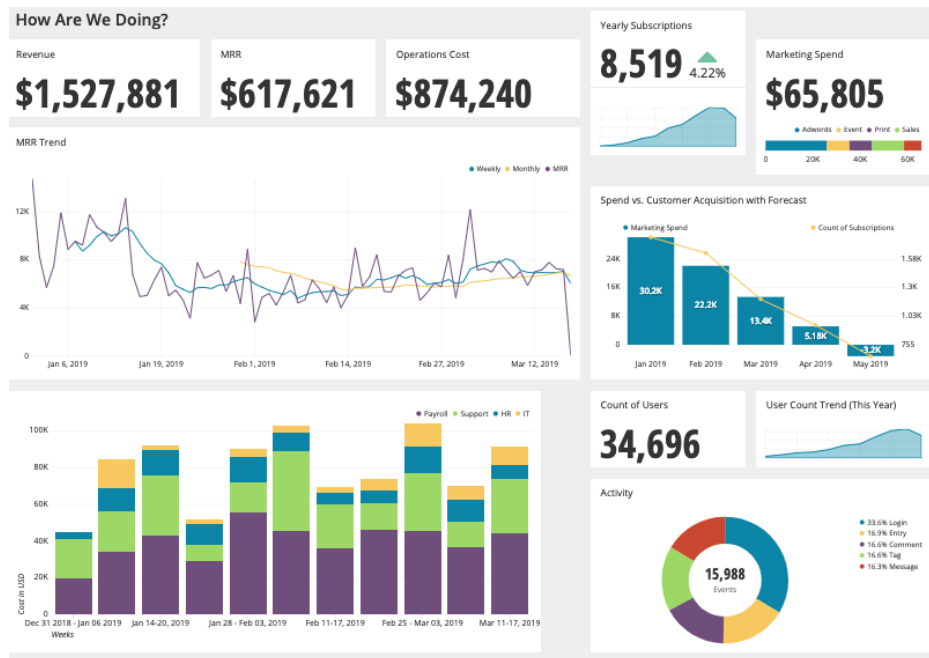
If data is hard to decipher, it won't be read. The more explanation or context needed to understand charts on a dashboard the less the dashboard is an effective intelligence tool. Remember, speed to insight is key. Squinting is challenging and time spent seeking out the author of a report for further explanation diminishes the impact of the dashboards goal.

Layout

Viewers in most countries read from left to right and top to bottom. Therefore the most important information should be top left and the least important information on the bottom right.



Visualizations should be aligned. Having a chart unaligned with the other visualizations will distract from the goal of presenting all of the information clearly to the viewer.



Since Yearly Subscriptions is out of alignment, it sticks out, and grabs the viewers attention and therefore may seem more important

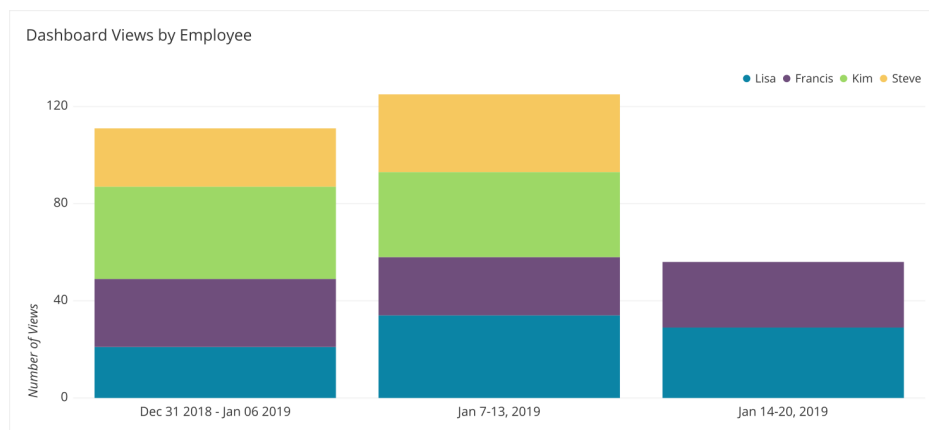
Empowering

Does a dashboard get used regularly and does it help people make decisions? These qualities are best evaluated after the dashboard is created by the end user/viewer of the dashboard.

- Do they view it regularly?
- Does it factor into their decision making?

Do they view it regularly?

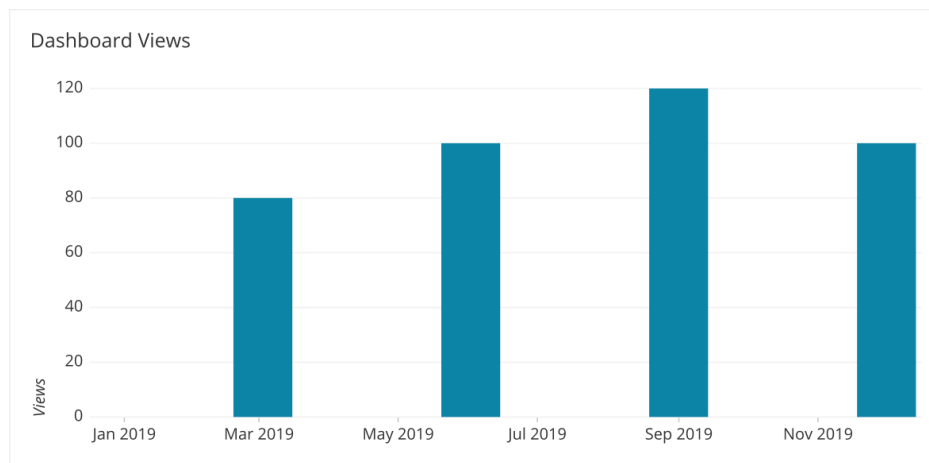
Most BI tools will provide you with a query log where you can track the number of views for each dashboard:



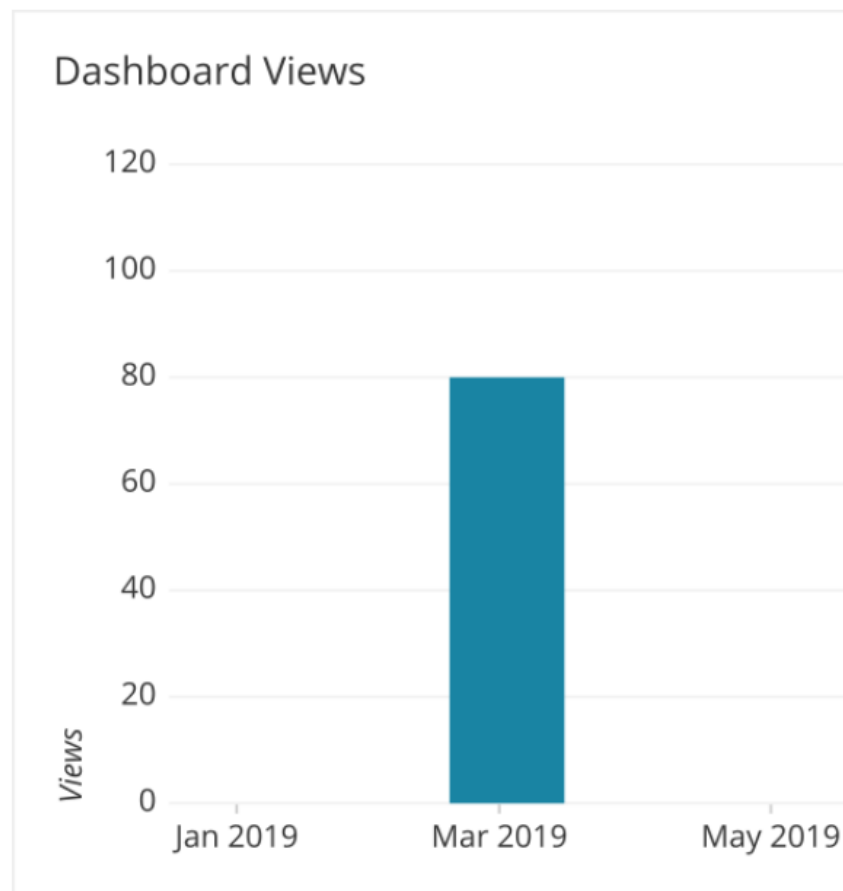
If you start to see a drop off in views, you should follow up with the individuals who are no longer viewing the dashboard. Asking the following questions can help:

- Is the dashboard useful to them?
- Is the dashboard missing some critical information?
- What sources are they currently using to access accurate data for making decisions?
- How can the dashboard be updated to better serve their needs??

Some dashboards may be viewed less frequently if they are set up for longer-term decisions. These gaps indicate when the dashboard is not being used. The regular spikes show when it is used and how useful it is.



In this case, the views might be related to quarterly planning or reporting, if you look at a shorter time frame it appears not useful.



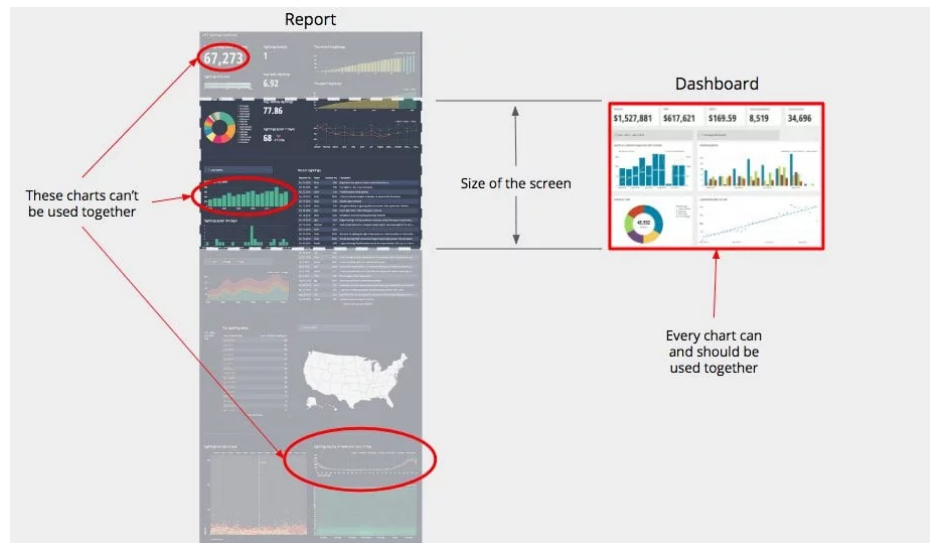
Does the data displayed on the dashboard factor into their decision making?

A simple way to check this is to ask the viewer, if the numbers on this dashboard went to 0 or if the numbers doubled would you do anything? If the answer is no, then the dashboard is probably not useful. If the answer is yes, then the dashboard is probably useful.

Succinct

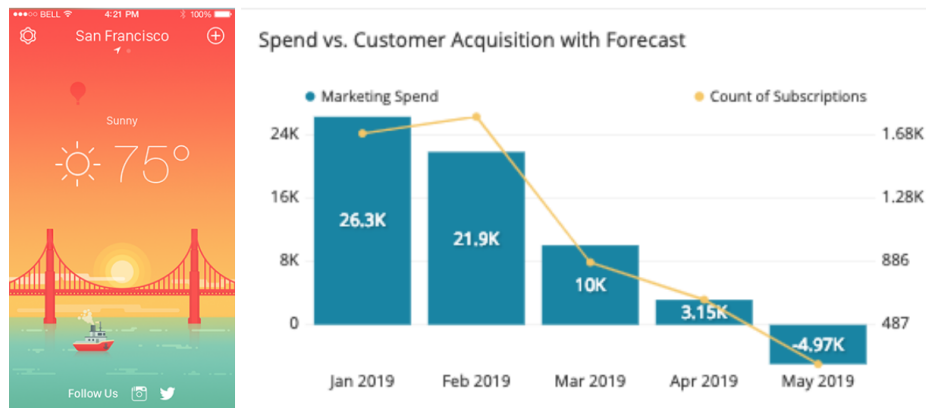
One of the main benefits of a dashboard is that it shows multiple data visualizations simultaneously which facilitates processing all of the information together. Due to people's limited working memory, needing to scroll to see other data visualizations prevents viewers from being able to compare the various visualizations side by side to reach significant conclusions. Scrolling becomes counterproductive.

In the image below we can see the report on the left it would be impossible to compare the charts that are circled at once since we would need to scroll to see each one.



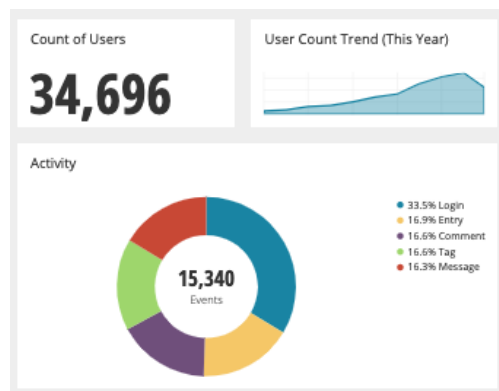
Having information on a dashboard hidden “below the fold” or below the bottom of the screen usually indicates that there is too much information on that dashboard. Ask yourself if the data is necessary or if it can be displayed in a smaller space without compromising its clarity.

It is easy to keep putting more and more charts on a dashboard that are useful, however not all useful charts are relevant to the purpose of a dashboard.



While a weather forecast is useful and a Spend vs Customer Acquisition with a Forecast chart is useful. Neither is relevant to each other. Keep your forecasts separate :-)

Data that is relevant to other pieces of information on the dashboard should be placed in close proximity on your dashboard



Looking back at our original dashboard, we can note the placement of relevant data succinctly which makes it is easy to evaluate the data displays as a group.

Summary- think ACES

- Accurate - If the data that is visualized is incorrect or the visualization biases how it is interpreted the dashboard becomes unusable
- Clear- clarity allows for speed of insight
- Empowering - will people access the dashboard regularly to make decisions (makes sure the dashboard delivered on its goal of supporting decisions)
- Succinct - keeping it brief and relevant to critical information makes it easy to evaluate all of the data simultaneously and make decisions

Dashboard Design Process

The dashboard design process is very similar to any design process however the output is a dashboard... obviously.

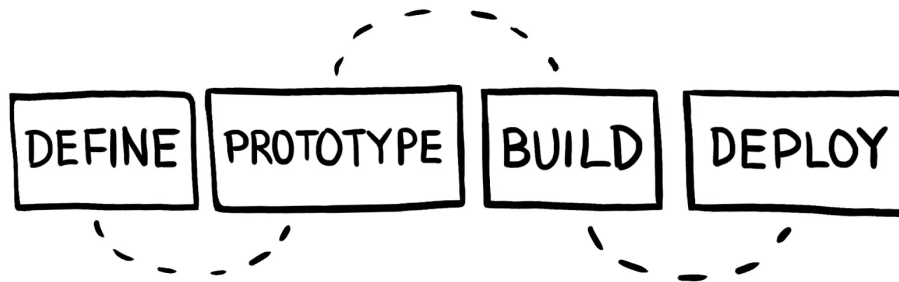
Why Process is Important

People get very attached to their solutions for a problem. The more they invest time and effort in their solution the stronger they believe in it regardless of the quality of the solution. This is why it is critical to avoid jumping right to solving the problem by building a dashboard. We don't want low quality dashboards.

Spend time understanding *why* we are building a dashboard and give ourselves *time* to explore multiple ideas before selecting what we will actually build.

The process starts with defining our stakeholders and what the metrics they care about. Then we prototype dashboards, get feedback, and iterate. After getting feedback that the prototype meets the goals of the project we need to find the actual data and build the dashboard. Finally we need to share the dashboard and maintain it so it becomes a useful tool for the audience.

We can summarize this process into 4 steps:



1. Define



This is the first and most important step. Having clarity about who this dashboard is for and what metrics matter to them is critical to creating a dashboard that will be used.

Stakeholders

There are 4 main stakeholders

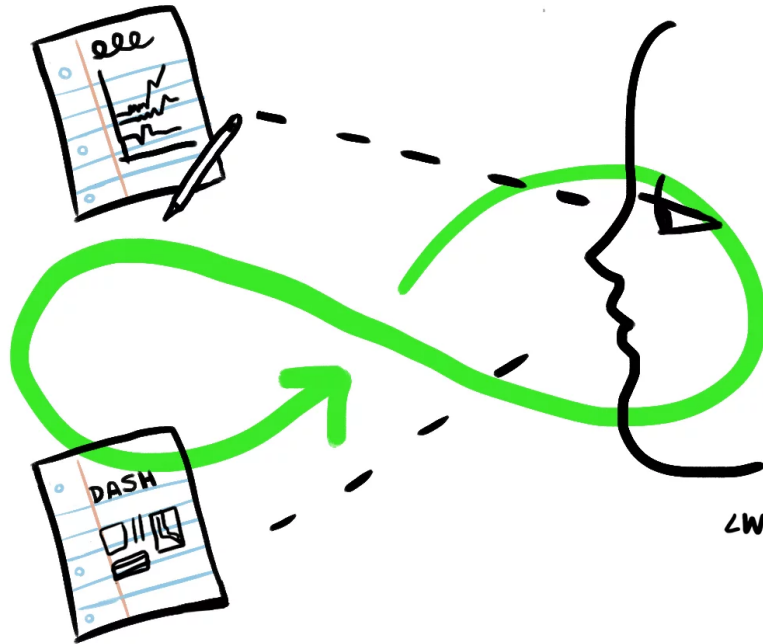
- The designer (you)
- The audience (who will be viewing this dashboard)
- The point person (the member of the audience who has the most experience)
- The Data Gatekeeper (member of the data team who will help with the database)

We will define in detail their responsibilities and where they factor into the process in the next chapter.

Metrics

You will work with the point person to go from vague goals to query-able metrics. This process involves a lot of back and forth to weed out interesting but ultimately not relevant metrics to mission critical data for decisions to be based on. We will go into further details about how to do this in [Determine the Metrics to Monitor](#).

2. Prototype



Once we have the metrics that we want to put in a dashboard we need to figure out how best to display them so that it is useful to the whole audience.

Visualizations

Use the visualization that best presents the metrics clearly and accurately. Even when sketching and prototyping graphs making the right visualization decisions here will improve the prototype and the feedback loop. We will cover when to use which chart in depth in [Find the Best Visualizations for Your Metrics](#).

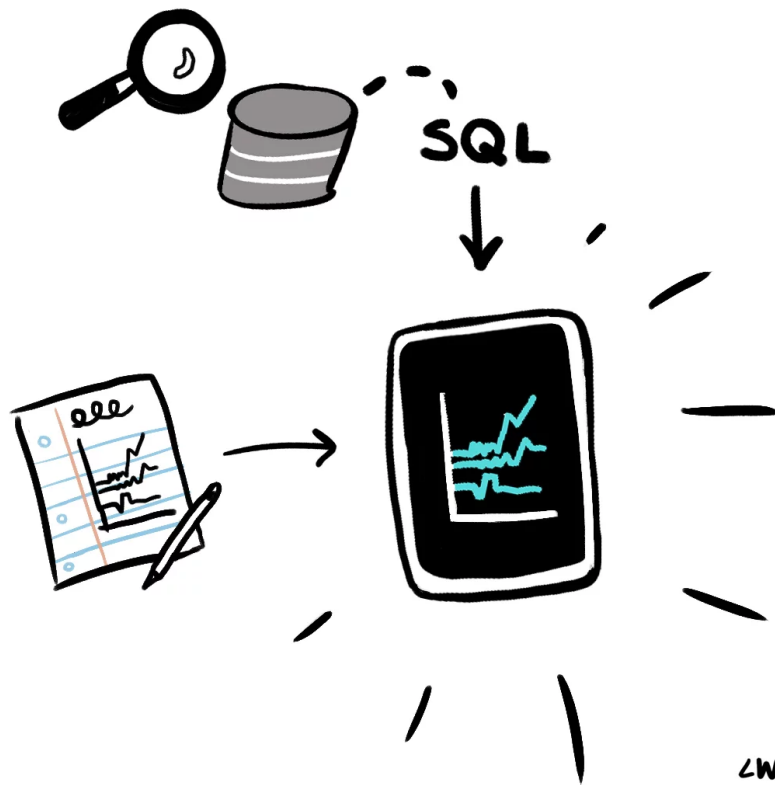
Dashboards

There are best practices for taking the visualizations and composing them together in a dashboard. In fact some composition choices might actually make you change what visualization you had selected as optimal before. We cover best practices for arranging dashboards in [Arranging Your Visualizations as a Dashboard](#).

Sketching and Iteration

At this stage it is recommended that the visualizations and dashboards be sketched out on paper or using a lo fi tool that is not connected to any real data. The reason for this is that it allows you to quickly throw out bad ideas without worrying about the time investment. It also allows you to focus on design instead of checking if the numbers are right. We talk through visualization and dashboard prototyping strategies in [Dashboard Prototyping and Feedback](#)

3. Build



Once we are satisfied with the prototype we have to create the dashboard using real data.

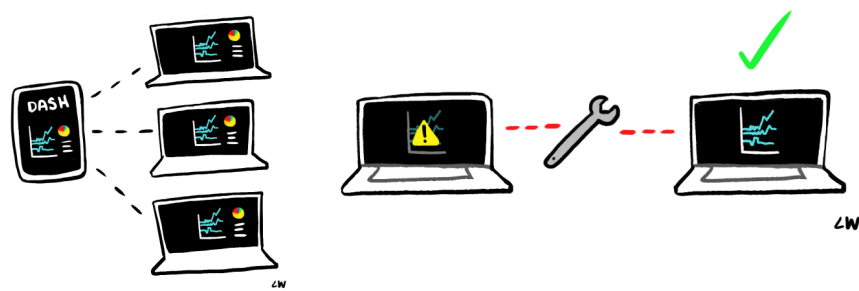
Find the Data

Many challenges can arise at this point. Where is the data stored? Is the data messy? Do we even have the data available? Working with the data team and the Data Gatekeeper is critical to navigating this step. We talk through common difficulties and how to overcome them in [Finding the Data That Builds Metrics](#)

Build Metrics/Dashboard

We need to create queries to power the metrics, create formulas, and transform the data into charts. Using a framework to log the metrics, formulas, and data sources makes creating queries much easier, we outline how to do this best in [Build the Metrics](#)

4. Deploy



Finally we have a fully functioning dashboard. Now we need to share it with the full audience. We should enhance the dashboard to make it more effective at scale and we need to make sure to maintain it as usage grows and changes.

Sharing

The audience will have varying levels of data literacy and context for the data presented in the dashboard. You need to verify you have enough context within the dashboard and that you provide enough training so that people can get insights out of it easily. We go over these techniques in [Sharing the Dashboard – Distribution Strategies](#)

Scaling

If the dashboard is useful the amount of views and total number of viewers is likely to grow. Adding links, interactivity, and documentation to a dashboard helps it accommodate more use cases and inspire other dashboard creators. Also as the number of views/viewers increases spending time optimizing queries becomes an important way of keeping the dashboard useful. We outline the steps in [Scaling Dashboards](#)

Maintenance

Datasources, tables, and fields change, dashboards need to change with it. Setting up scheduled times to review dashboards is critical to keeping them relevant and functional. Providing a way for the audience to alert you about issues will allow you to make informed improvements to the dashboard. We cover maintenance in [Making Sure Your Dashboard Always Gets Better](#)

Define

Identifying Key Roles

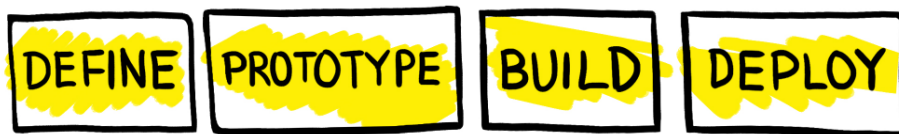


In a dashboard design project, multiple people are involved in making it a success. Key stakeholders need to be defined from the start to facilitate optimal collaboration and communication during the appropriate steps of the process. There are 4 key stakeholders with corresponding responsibilities in a typical dashboard project.

These roles may be fulfilled by the same person but typically they are not.

Key Stakeholders

1. Designer

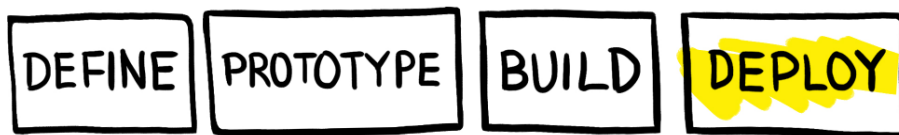


This is the person that is responsible for creating the dashboard, and managing the project. As you are the one reading this book this is likely you. Since you will be involved in all steps you should be in charge of coordinating all efforts of the project.

Designer responsibilities:

- Define the stakeholders
- Metric calculation
- SQL queries
- Chart determination
- Dashboard build
- Point Person approval

2. Audience



This stakeholder group is the most passive in the creation of the dashboard, but the most active once the dashboard is finalized.

Audience Responsibilities:

- Provide feedback on whether or not the dashboard is useful

There are no direct responsibilities for this role during the design and build phase, however, all design decisions should take into account how the consumer will interact with the dashboard. Pretty much every decision that is made in this project should be made with the Consumer in mind. Their product or business knowledge, business role, and any cultural or other sensitivities NEED to be considered when completing almost every step of the dashboard project.

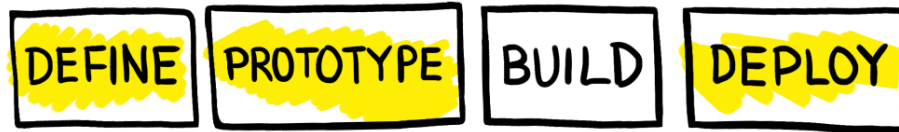
Questions to keep in mind about Consumers:

- What devices will they use to view this dashboard?
- How much context is needed for each metric?
- How often will they be referring to this dashboard?

These questions will help guide your design decisions so that the dashboard is most effective for the consumers.

However there is one audience member who is very involved in the process, we will call this individual the Point Person.

3. Point Person

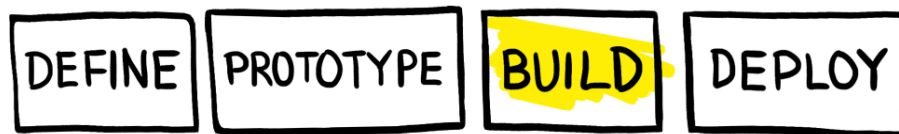


The Point Person is the one that is requesting the dashboard. This is the person who has the business need for creating the dashboard, such as wanting to understand the marketing funnel or how users are engaging with the product over time. They should be involved in all of the decision points since they have the most context to ensure the dashboard is useful.

Point Person's responsibilities:

- Provide the decisions this dashboard will inform
- Identify the key metrics
- Identify the audience of this dashboard
- Provide approval on the dashboard design

4. Data Gatekeeper



This is a member of the data team that understands the data source, schema, and any other library that will be used to create the metrics. This may also be you or perhaps not.

Data Gatekeeper's main responsibilities include:

- Setting up permissions and access to the data
- Helping others understand the data source schema
- Helping with SQL queries

Figuring out which columns to use in a visualization and in a dashboard can be challenging since many databases are not well documented and the data is not modeled appropriately. Data Gatekeepers can make this much easier since they are familiar with the schema and can grant you permission to tables that you may not have access to. They may also write the SQL queries themselves or help you optimize queries that you have written.

Summary

- Creating a great dashboard is a team effort, and all roles are vital in its completion.
- Designers coordinate the project and create the dashboard
- Consumers are the audience, and the design should meet their needs
- Point People define the central decisions that a dashboard needs to support
- Data Gatekeepers help find the relevant data in the database

Determine the Metrics to Monitor



After defining the stakeholders involved we need to get clear about what decisions this dashboard is going to help people with. The point of a dashboard is to help people find insights and make decisions. The Point Person will help you determine this since they are the main audience member who will be using the dashboard to make decisions.

Often times the Point Person will have really vague overall goal for wanting a dashboard. It is your job to help them refine their vision to a few key decisions that need to be made and then finding metrics that will help inform those decisions.

- Understanding the Point Person
- Define the Decisions
- Define the Metrics

Understanding the Point Person

Overall Goal

While there is a lot written about strategies for defining goals, they all boil down to the same thing. State a specific result that you want to achieve. The more specific the easier it is to act on. Have the Point Person write out what specific result they are hoping to achieve by having a dashboard.

Ask them: *What does success look like?*

The more specific they are the easier the rest of the process is since the dashboard can be measured against that specific result.

Motivation

Understanding what they want out of the dashboard is a great start but we want to dig deeper to understand their motivation, what is the pain the dashboard will solve for them. We can understand this by learning how they are currently solving their problem. Dig into their stated goal using a customer development framework such as:

What is the hardest thing about achieving [stated goal]?

Why is that hard?

Can you tell me about a time when you last tried achieving that goal?

How did you overcome the hardest parts?

What don't you like about your solution?

We will learn a lot from the Point Person's answers. If it is not hard for them to achieve their goal a new dashboard is unlikely to be valuable to them. Understanding why achieving their goal is hard gives insight into why a dashboard might be the solution they need. Hearing a story about their struggle in achieving their goal will give you insight into the full process involved. If they have not tried to overcome the hard parts they likely do not care that deeply about the goal. If they did try to overcome the hard parts then you might be able to borrow from their solution in your dashboard. Understanding what they didn't like about their solution will make sure you do not repeat the same mistake in your dashboard.

After getting perspective on the goal and motivation of the Point Person we need to breakdown all of the decisions along the way to achieving their goal that need data to support them.

Define the Decisions

Best Case Scenario

Start with ideal scenario, and work backwards. Have the Point Person imagine that they have achieved their goal and then think backwards.

- How did this happen?
- What decisions did they make in order to achieve their goal?
- What data did they have to make those decisions on?

This exercise eliminates any technical constraints and captures the aspirational decisions they wish a dashboard could support.

Worst Case Scenario

Explore the worst scenarios, and work backwards. Have the Point Person imagine completely failing at achieving their goal.

- How did this happen?
- What bad decisions lead to this outcome?
- What decisions were made with a lack of data?

This exercise identifies decisions that definitely need data to support them. It focuses on identifying what could go wrong so it will likely surface more practical decisions.

Prioritize

Take the list of aspirational and practical decisions and rank order them with the Point Person. While there is no hard number for the number of decisions you want to support with a dashboard the fewer the better. It creates a focused use case and makes it easier to design the dashboard. After determining the decisions for this dashboard we need to see how we could inform those decisions with metrics.

Define the Metrics

After we know what decisions are important we need to figure out which metrics would best support those decisions. Below is a spreadsheet that outlines all the necessary pieces that you will need to specify to create a query around a metric. We will refer to this as the Metric Spreadsheet.

Work with the Point Person to write out how they want to aggregate, group by, and filter the metrics. If you are skilled at extracting metric formulas from business people this may not be necessary, but if you want a clear way to document their metric definitions we suggest using a Metric Spreadsheet. Either way, this book will use the Metric Spreadsheet to keep track and reference what metrics are important to this dashboard design process.

Metric Spreadsheet

Metric	Aggregation		Grouping		Filtering
	Formula	Data Source	Content	Data Source	

This Metric Spreadsheet will help you track what the Point Person wants and help you build the final visualizations quickly.

There are two approaches that can assist with figuring out the proper metrics that will support their decision making. Visual and conversational. They can be helpful independently or used together.

Visual Method

Some people who are requesting a dashboard will struggle to articulate what they need. An easy hack for this is to have them fill out what a dashboard would look like that would help them make decisions. Here is a sample template:

<div><div>Title</div><div>DRAW GRAPH HERE</div></div>	<div><div>Title</div><div>DRAW GRAPH HERE</div></div>	<div><div>Title</div><div>DRAW GRAPH HERE</div></div>
<div><div>Title</div><div>DRAW GRAPH HERE</div></div>		<div><div>Title</div><div>DRAW GRAPH HERE</div></div>

After they fill this out it is easier to talk with them in more detail about the underlying metrics in the visualizations they drew. In the “Bringing it all together” section further down we will discuss how to dig in to each of these visualizations.

Conversational Method

After determining their prioritized list of decisions that they need to make, clarify what metrics would inform those decisions with the following questions:

- What data would help inform that decision?

As you work through each decision with the Point Person, there will likely be multiple types of data they wish they had for each decision.

You should also directly prompt them at each question by asking “What other data might be useful” or “How else could we improve this decision’s quality?” The data they want to see will become the metrics that you will be visualizing for them in the dashboard. Before we get started building them we want to make sure we understand exactly how these metrics are aggregated, grouped, and filtered.

Bringing it all together

After determining the set of metrics the Point Person is interested in from their drawings or through conversation, get specific about each one.

- How is this metric calculated?
 - Write out formula behind each one. For example: LTV is Average Revenue Per Customer / Customer Churn Rate
- How do you want to group the data?
 - Category, Timeframe, etc
- How do you want to filter the data?
 - Group
 - Timeframe
 - Specific Criteria

It is valuable to push the Point Person to think about the question in different ways. Go through common aggregations(SUM, COUNT, AVG, MAX, MIN, etc) and groupings (Time, Category, etc) to see if they find any value in them. Looking at a metric in multiple ways can help people develop better insights around that data. Should they look at the average or the median, should they group by month or by year.

After they clarify all these aspects of the metrics they are interested in writing it down so you can easily calculate them in your BI tool. You could also use this information in the Metric Spreadsheet to log the information.

We will use the Metric Spreadsheet to keep track, the questions above map to directly to it's columns:

Metric	How is this metric calculated? ↓ V		How do you want to group the data? ↓ V		How do you want to filter the data? ↓ V
	Aggregation		Grouping		Filtering
	Formula	Data Source	Content	Data Source	
Operations Cost	SUM(Cost)		Total Department Month		Do not include marketing costs
Revenue	SUM(Revenue)		Total Month		
Subscriptions	Count(Accounts)		Total Month		Do not include free trials

These metrics should inform the decisions that need to be made by the Point Person to achieve their goal. Before we start prototyping these metrics in the next few chapters we will review best practices for visualization selection and arranging a dashboard.

Summary

- The first step in building a great dashboard is understanding the central premise or purpose for building the dashboard.
- This understanding can be obtained by interviewing the Point Person of the dashboard and first finding out the question they want to have answered.
- Then you can work with the Point Person to define the metrics that are involved in answering the questions that will guide their decision making.

Prototype

Find the Best Visualizations for Your Metrics

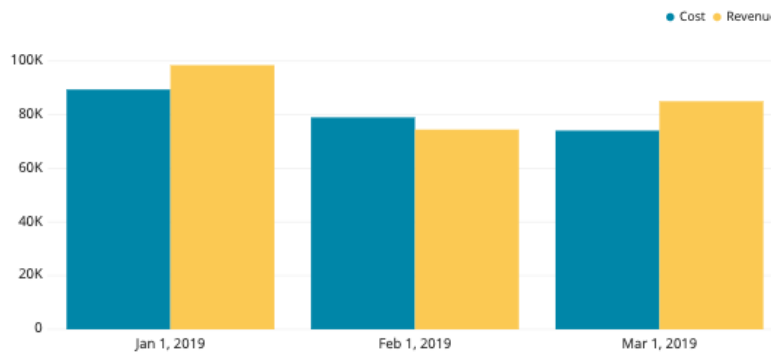
Selecting the best chart for your metrics is not always a straightforward process. Certain visualizations do not represent some datasets well, and certain visualizations can not represent some datasets at all. Sometimes you just need a table, single value, or just show some text. Knowing the difference will help you design the most useful dashboards.

Advantages of data visualization

Graphs help people recognize patterns faster than looking through a table with numbers in it. For example, take a look at the table below..

Date	Revenue	Cost
1-1-2019	98473	89403
2-1-2019	74459	79004
3-1-2019	85013	74093

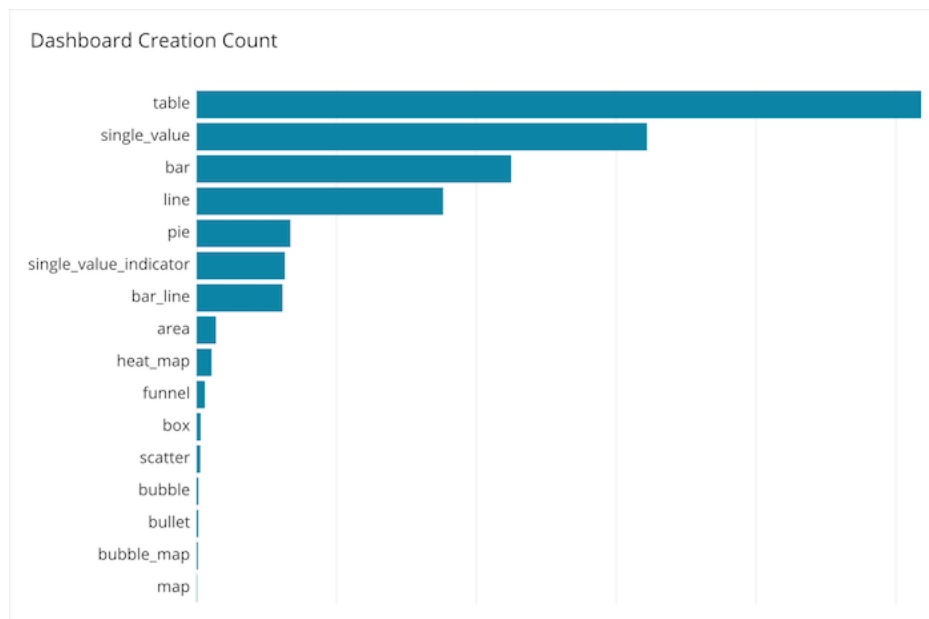
Now take a look at this chart...



Data visualization is a general term that describes any effort to help people understand the significance of data by placing it in a visual context. Patterns, trends and correlations that might go undetected in text-based data can be exposed and recognized easier with data visualization software.

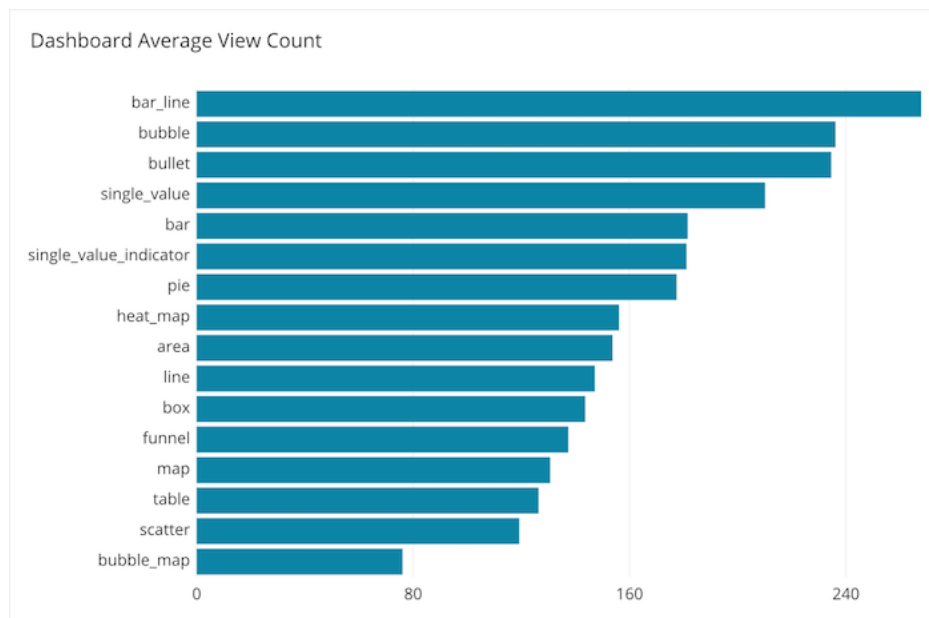
Most Common Visualizations

After studying 90,000 dashboards at Chartio we found most data is displayed in a handful of the chart options.



While these may not be the most optimized they are most often created. People create table views of their data, single values, bar charts and line charts.

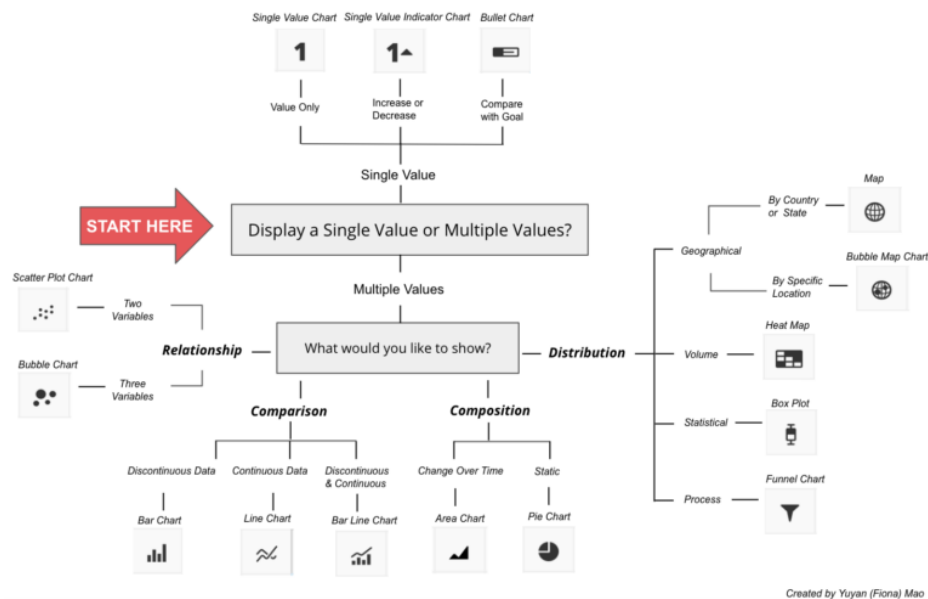
However when we look at what visualizations are on dashboards that get the highest average views we get a different ranking.



Bar Line, bubble, bullet, single value, and bar charts are the most often viewed. Consider these options before going into the more specialized types of visualizations.

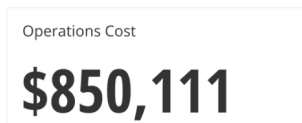
Selecting Visualizations

We also created a list of when each chart type is optimal to use for viewers interpreting the data correctly. We have created a decision tree to help you choose the most effective chart for your data. You can use this flowchart to select your visualizations. Please download this, print it out, and put it on the wall near your desk.



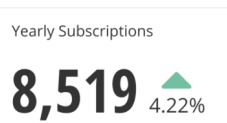
Single Value

Single Value Chart



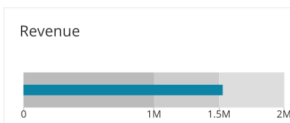
Used to draw quick attention to high-level metrics.

Single Value Indicator Chart



Used to draw quick attention to high-level metrics with an indication of movement.

Bullet Chart



Used to draw quick attention to high-level metrics in relation to a goal.

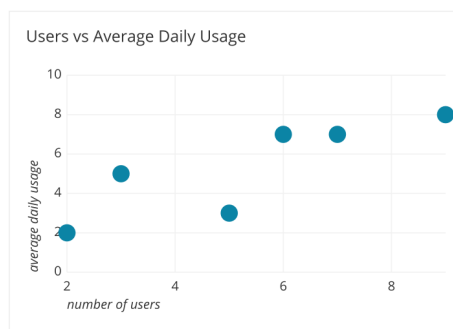
Multiple Values

To display multiple values there are 4 common categories:

- Relationship: how multiple independent variables relate to each other.
- Comparison: how two or more sets of data compare with each other.
- Composition: how a set of data is made up of smaller divisions.
- Distribution: how different sets of data are spread over a population or other distribution.

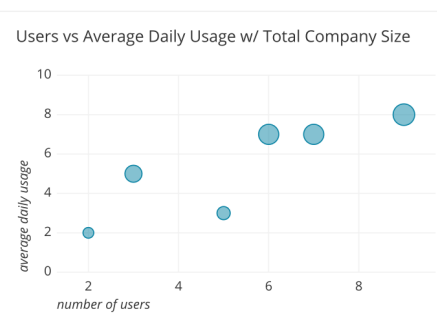
Relationship

Scatter Plot



Shows the relationship between two quantitative variables.

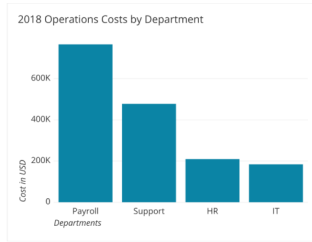
Bubble Plot



Shows the relationship between three quantitative variables.

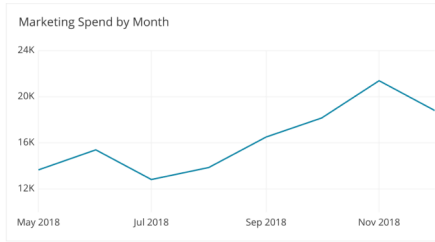
Comparison

Bar Chart



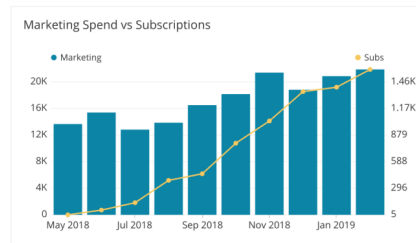
Used to show discontinuous data

Line Chart



Used to show continuous data

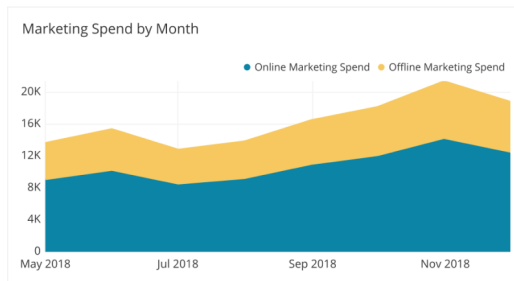
Bar Line Chart



Used to show discontinuous and continuous data

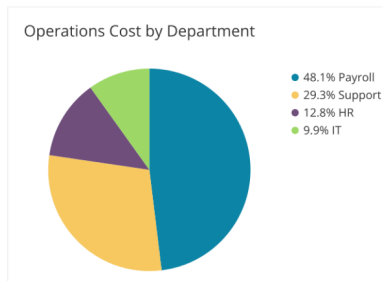
Composition

Area Chart



Shows the composition of a continuous variable over time

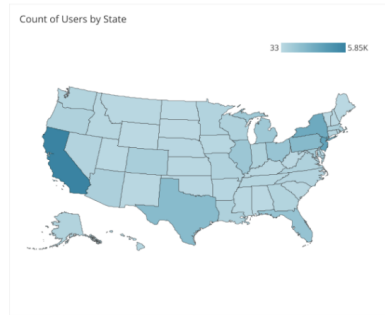
Pie Chart



Shows the composition of a variable in a single point in time.

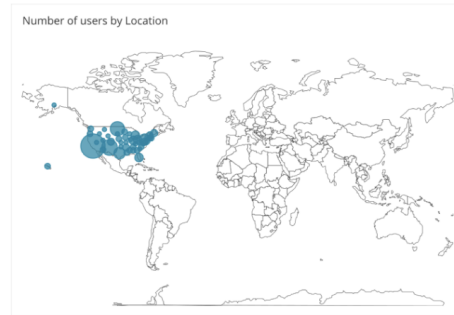
Distribution

Map



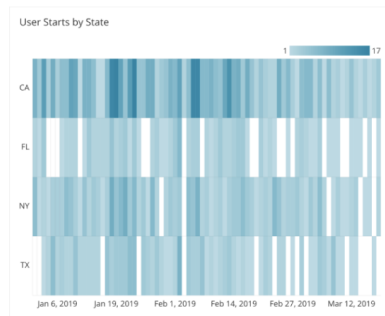
Used to show how data is distributed by geo-political boundaries

Bubble Map



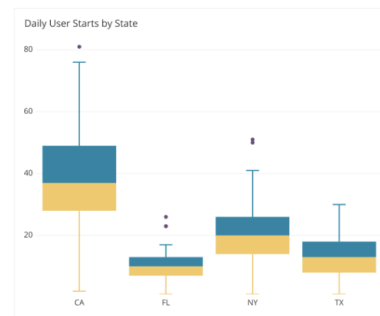
Used to show how data is distributed by precise location

Heat Map



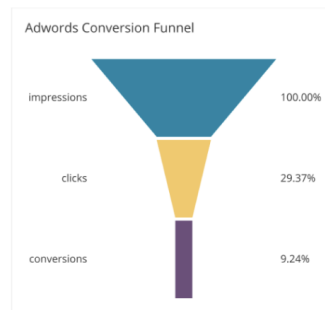
Used to identify temporal patterns (days, weeks, months) or areas of concentration.

Box Plot



Used to show single or multiple distributions in a compact view.

Funnel Chart



Used to show where drop off happens in a process

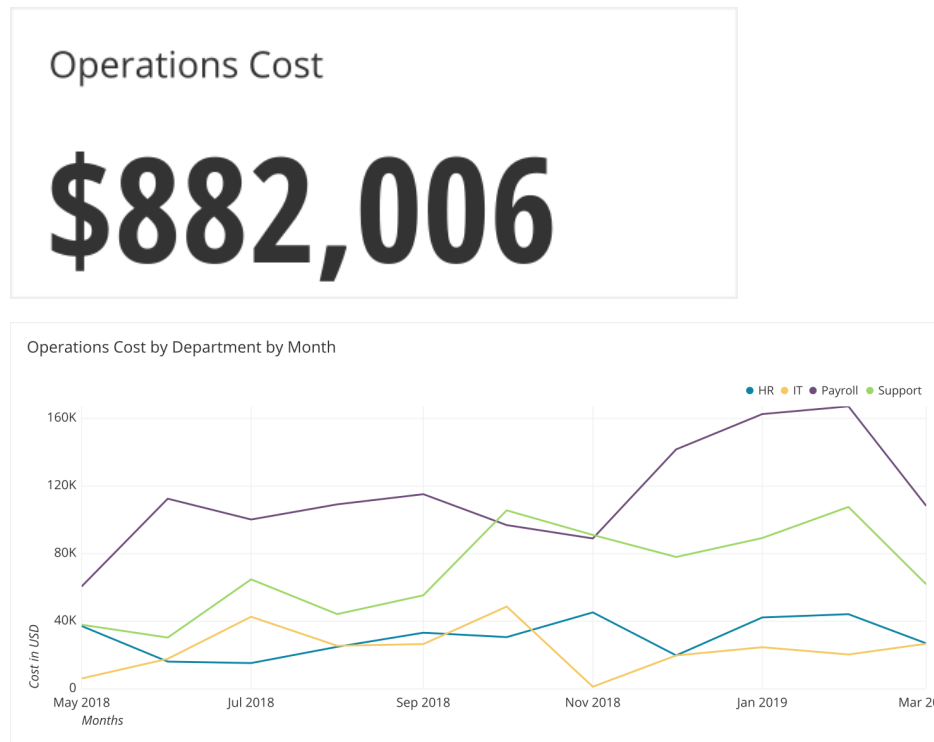
Examples

Using the Flow Chart we can go through the decision making process for each of the metrics we created in the previous chapter. Review the formula and grouping content in the Metric Spreadsheet and lets create visualizations for each one.

Metric	Aggregation		Grouping		Filtering
	Formula	Data Source	Content	Data Source	
Operations Cost	SUM(Cost)		Total Department Month		Do not include marketing costs
Revenue	SUM(Revenue)		Total Month		
Subscriptions	Count(Accounts)		Total Month		Do not include free trials

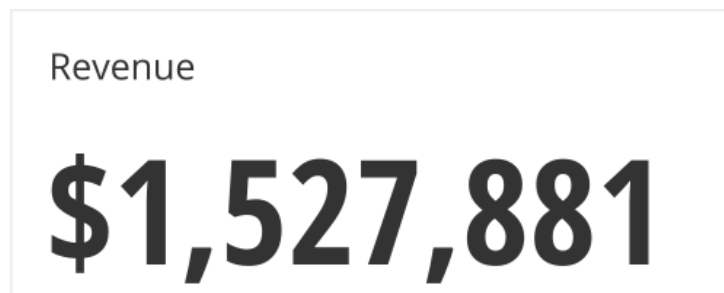
Operations Cost

There are multiple ways we need to aggregate operational cost. We need to calculate the total and we need to calculate the total grouped by department and month. For the total a single value chart is appropriate. For the by department by month aggregation we should use a line chart



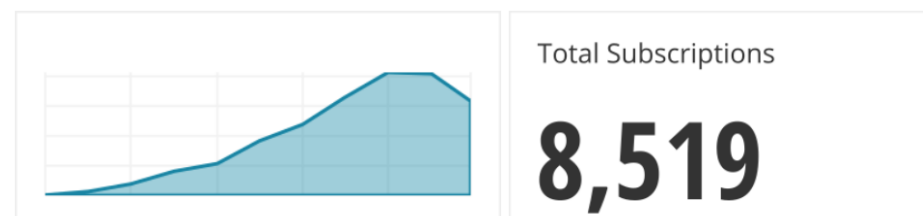
Revenue

As we identified in the table earlier, the result we need is a single value. As the Point Person didn't leave a comment about wanting to compare it to previous periods or to a goal, the "Single Value" chart is the best way to represent the figure.



Subscriptions

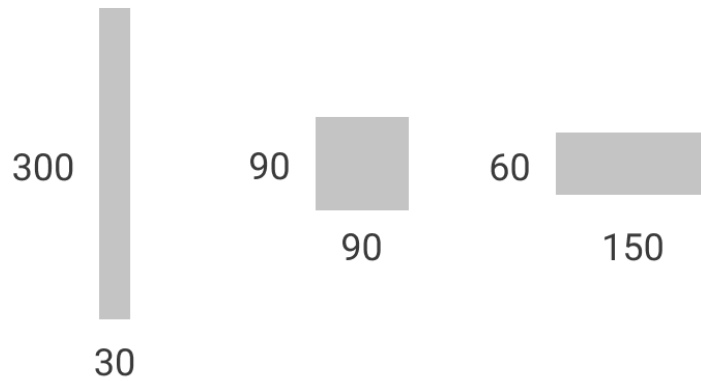
We are provided with one number, but we can add a sparkline (a small line graph without axes) to give us an idea of the historical context of how the number has moved.



Things NOT to do

3 Dimensions

People already struggle with comparing 2D areas, for instance all of these have the same area.



This is also why many people recommend caution using pie charts or area charts since they leverage 2D space. They can be used but only when there are very few categories.

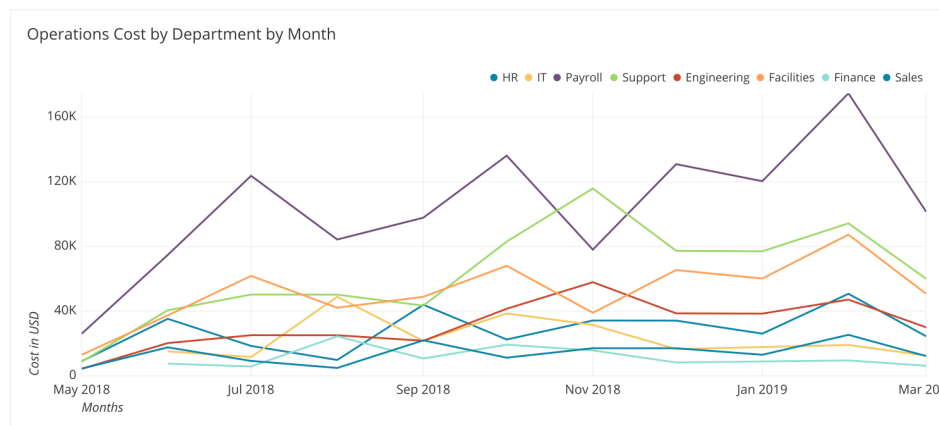
People are also not effective at comparing 3D volume, all of these shapes (except for the green L at the top center) have the same volume (made up of 4 cubes of the same size).



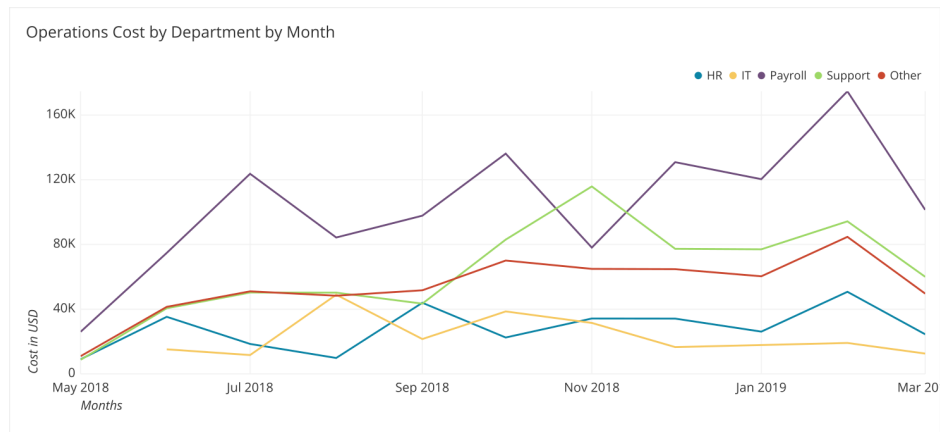
In addition it typically requires much more space to convey 3D information than showing it in 2D. Stick with simple 2D visualizations.

Too many categories

Once there are more than 5-7 categories it can be difficult to understand the graph quickly.



Consider limiting the number of categories shown in a visualization or try lumping together smaller categories into an “Other” bucket.



Summary

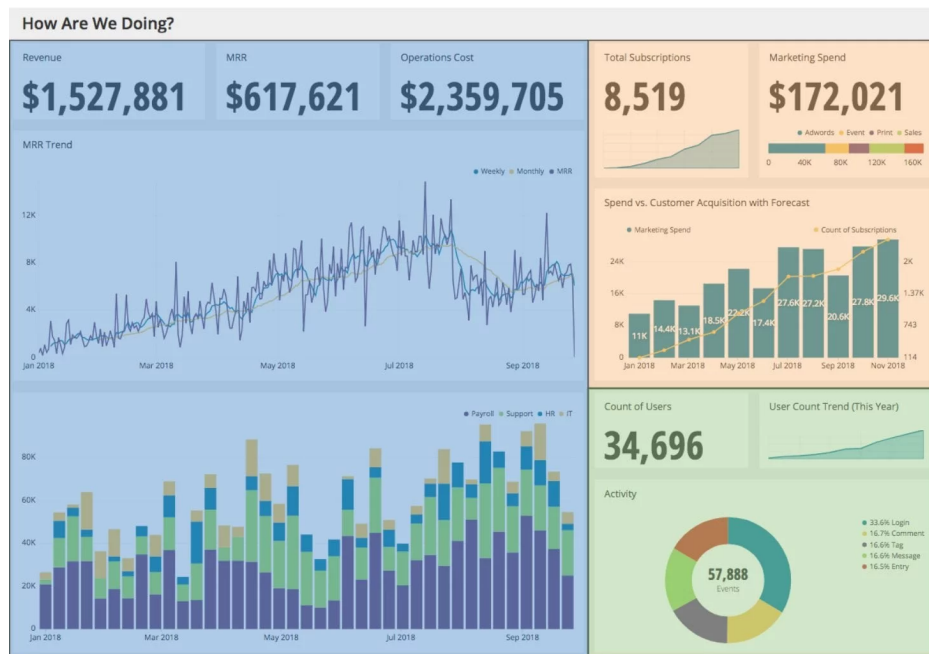
- Chart decisions should be made based on the data, using the flowchart can help you make the best decision
- If you are displaying multiple values consider whether you are trying to show relationship, comparison, composition, or distribution between the value to help you pick the most appropriate chart.

Arranging Your Charts as a Dashboard

Arranging the charts is where the science of a dashboard project starts to cede some of its control to art. There are many things to consider when arranging and sizing charts and selecting fonts and colors. How visualizations are composed in a dashboard can either facilitate or hinder the decision making process.

Alignment & Grouping

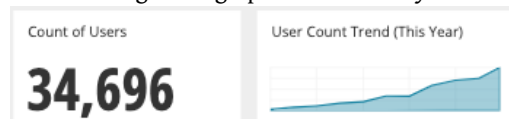
In the dashboard that we have been working to create with this book, we ended up with thirteen individual charts, monitoring eight different metrics. Some metrics were monitored twice so the current status, as well as a trend analysis over time, could be analyzed. When we started to create the dashboard we ended up with thirteen charts in three zones. By grouping charts and metrics that are closely associated with each other it gives us a larger picture.



Here we can see revenue and operations metrics shaded in blue, marketing metrics shaded in orange and user metrics shaded in green. Group charts that are about similar topics together.

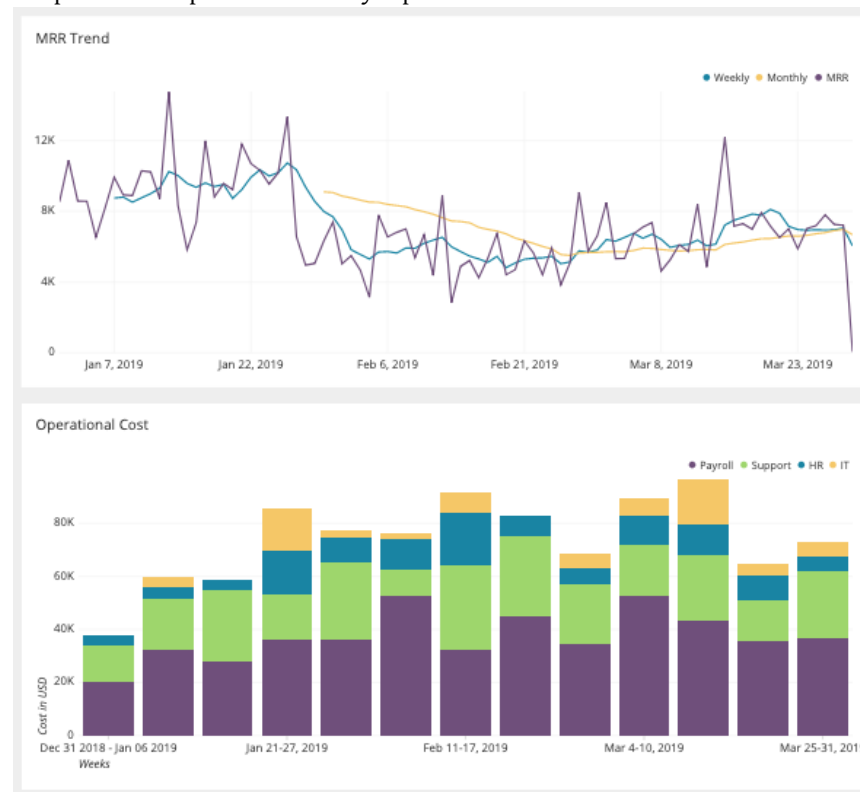
Alignment Quick Tips

- If a specific metric is displayed in two ways, these should be right next to each other. For example here we can see both the current Count of Users, and we can see how they have been trending in the graph next to it this year.



- Line charts with the same time frame should be stacked vertically to facilitate comparisons. For example below we can see how our MRR (Monthly Recurring Revenue)

compares to our operational costs by department.



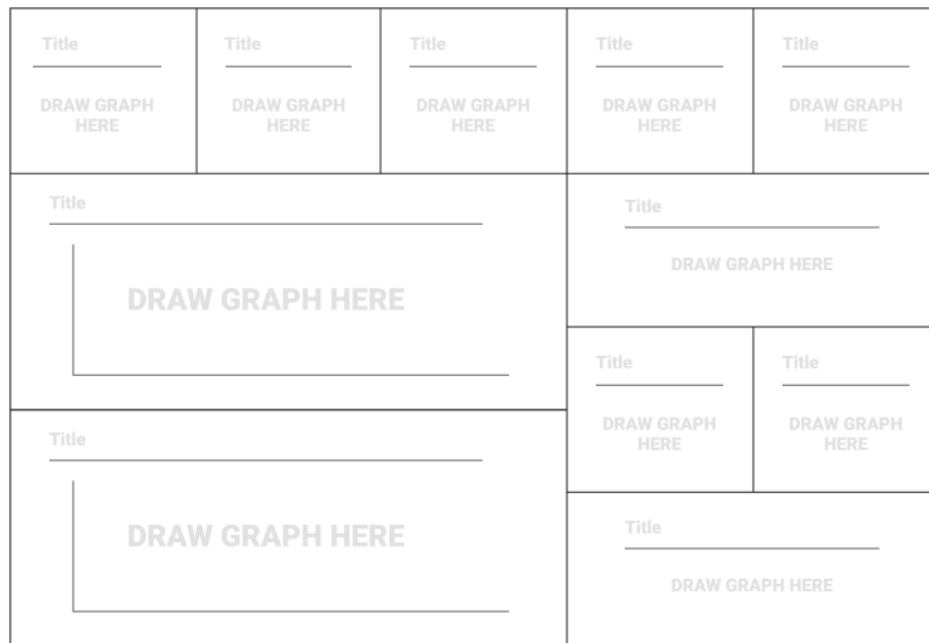
- Put items next to each other with no separation to use the same title for both visualizations. Here you can see that there does not appear to be any separation between these two visualizations. This groups the data and makes it clear they are describing the



same thing.

Examine Layouts

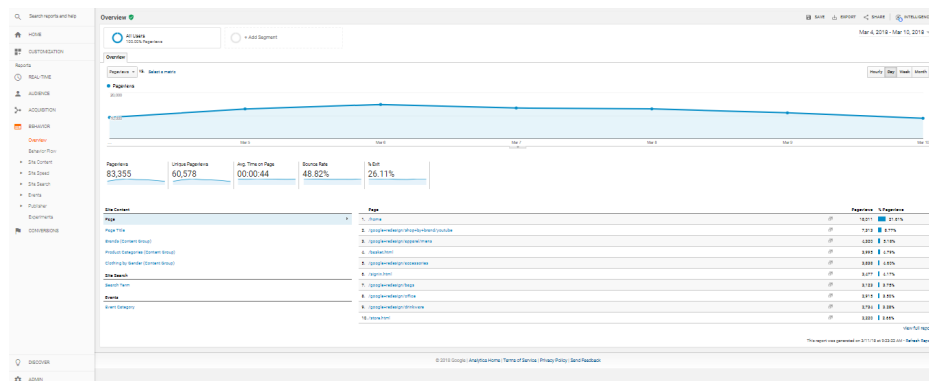
Determine the aspect ratio of the device (iphone, laptop, tv, etc) the dashboard you are building will ultimately be delivered on by consulting with the Point Person. View our templates to get ideas about how you could compose visualizations together.



View more here: [Dashboard Templates in Figma](#)

Also now is a good time to use google and apps that you use with a dashboard interface to get inspired about what combinations work.

Here is a great layout used by Google Analytics:

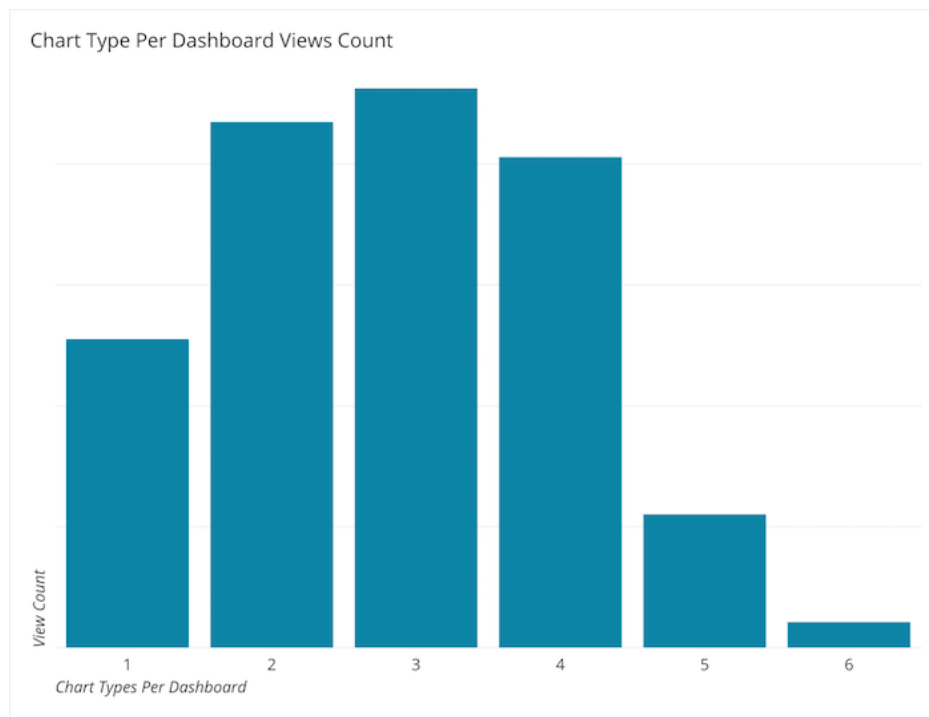


<https://medium.freecodecamp.org/how-and-why-to-get-started-with-google-analytics-153dc35b7812>

While being innovative in dashboard design can challenge the way we look at things in a positive way it is also likely to confuse the audience. Stick with simple arrangements that don't violate the ACES dashboard design qualities in our [previous chapter](#).

Limited Variety

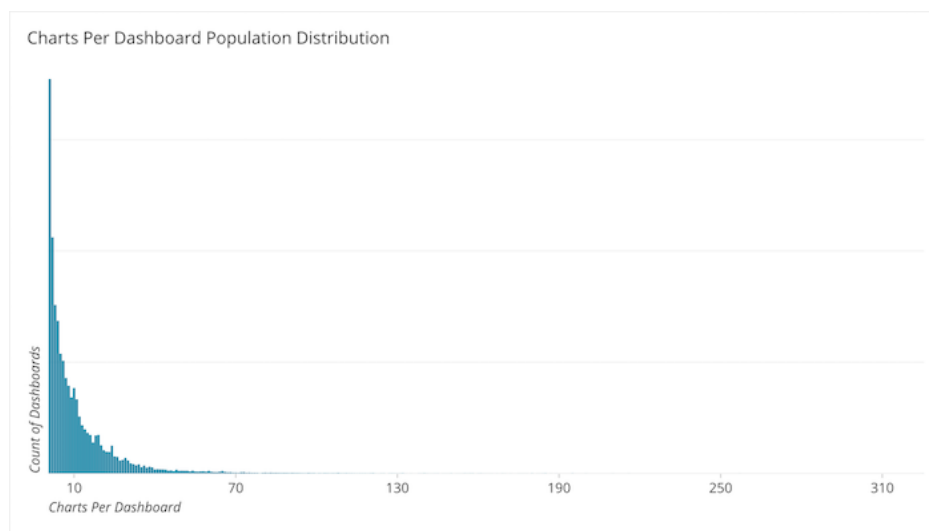
At Chartio we have also noticed that dashboards with multiple chart types get more views to a point. Viewership peaks when three different types of charts are used, and then it drops off.



Avoid Too Much Information (TMI)

In addition to not including too many cool BI Tool features in your dashboard, remember to not include too many visualizations in your dashboard either. There will be many opportunities to overload the dashboard with too much information. Information overload deteriorates the power of the dashboard.

At Chartio we took a look at the number of visualizations per dashboard and found that it is heavily skewed, the vast majority only use a few charts.



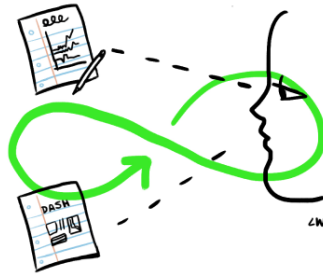
If information needs to be viewed as is and aggregated at a more granular level, you need a second dashboard containing this has a more drilled down view. Don't be afraid to link out to other dashboards. This preserves the decisions and goals of the dashboard at hand and let's people explore the data in more detail if they wish.

Summary

- Align visualizations to facilitate comparisons
- Group visualizations to facilitate insight
- Familiarize yourself with various dashboard layouts to help develop useful arrangements for your own dashboard.
- Be wary of flashy dashboards arrangements made possible in BI tools

- Do not overcrowd the dashboard with too many visualizations, link out to provide further investigation where appropriate

Dashboard Prototyping and Feedback



Review

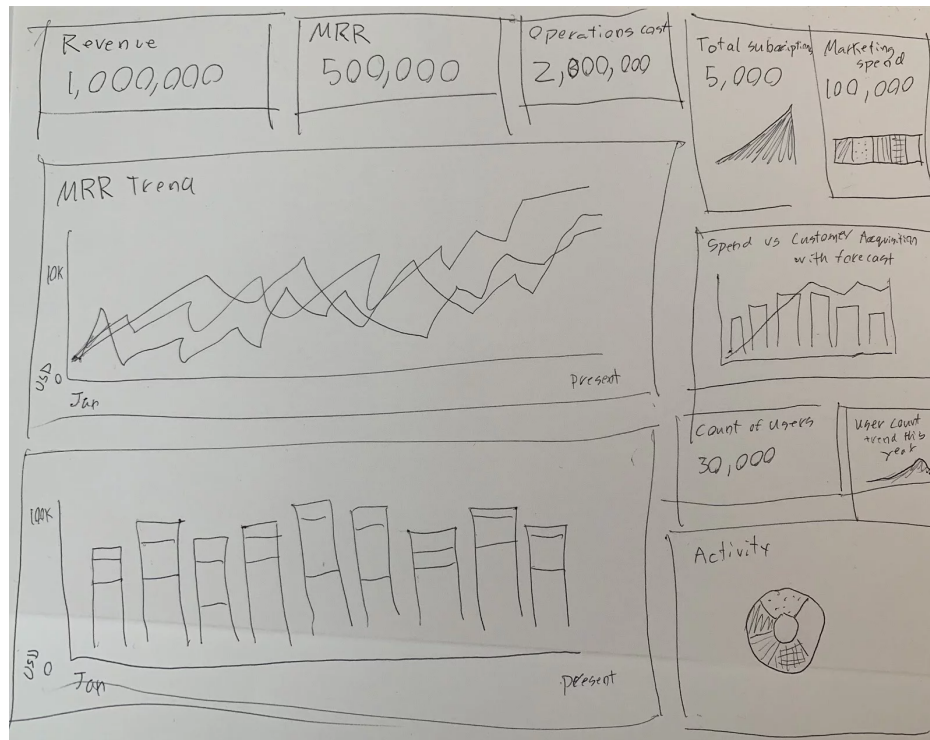
First gather your documents stating the goal, decisions, and metrics in one place. Do a quick review to make sure you understand all the requirements of this dashboard.

Now it's time to put pen to paper. In a typical design process the prototyping stage involves heavy divergent thinking to explore a wide range of solutions. In a dashboard design process the output is pretty heavily constrained to a few chart types. This means that you can start sketching pretty accurate designs right away.

Sketch

Start with individual metrics and draw them as visualizations. Then try drawing multiple visualizations together. Keep asking yourself

- What does the audience want to be able to do?
- What does the audience need to know in order to do it?



It can be useful to role play this scenario in your head or with others to understand what visualizations will be valuable.

As you sketch out different combinations of visualizations for the dashboard try to limit the number you use in each attempt to as few as possible. It can be helpful to adopt a template for structuring parts of the dashboard.

We have tried a few that we have found useful such as

- Indicator Chart -> table with relevant details
- High Level number this year -> line graph showing daily information
- Overall line chart -> same chart broken out by group

Feedback

You definitely want the Point Person's feedback as to whether they think it would support their decisions. This is also a good time to bring in other members of the audience for feedback. There are a few easy tricks to helping the Point Person and audience give you helpful feedback.

Ask the following questions for each visualization:

- Do I need to know this?
- What else do I want to know about this metric?
- Is the conclusion I can draw from the visualization obvious?

Ask the following questions for the whole dashboard:

- Does this dashboard support the decisions and goals of this project?
- Which visualization grabs your eye first? Should it?
- Should any of these visualizations be grouped together?
- Is anything missing from the dashboard?

Iteration

Be open to feedback. The goal is to create a useful dashboard for the audience, not to convince people your design is great. You will likely need to do more than one round of prototyping, feedback, and iteration. After you get positive signals from the Point Person that if they had this dashboard it would help support their decisions to reach their goal you are ready to bring the chart to life.

Build

Finding the Data That Builds Metrics



From the previous chapters most of the ambiguity of what is going on the dashboard should have been addressed:

- Why - why you are building this dashboard
- Who - who you will be working with in the dashboard creation process
- What - what metrics will support decisions to achieve the dashboards goal

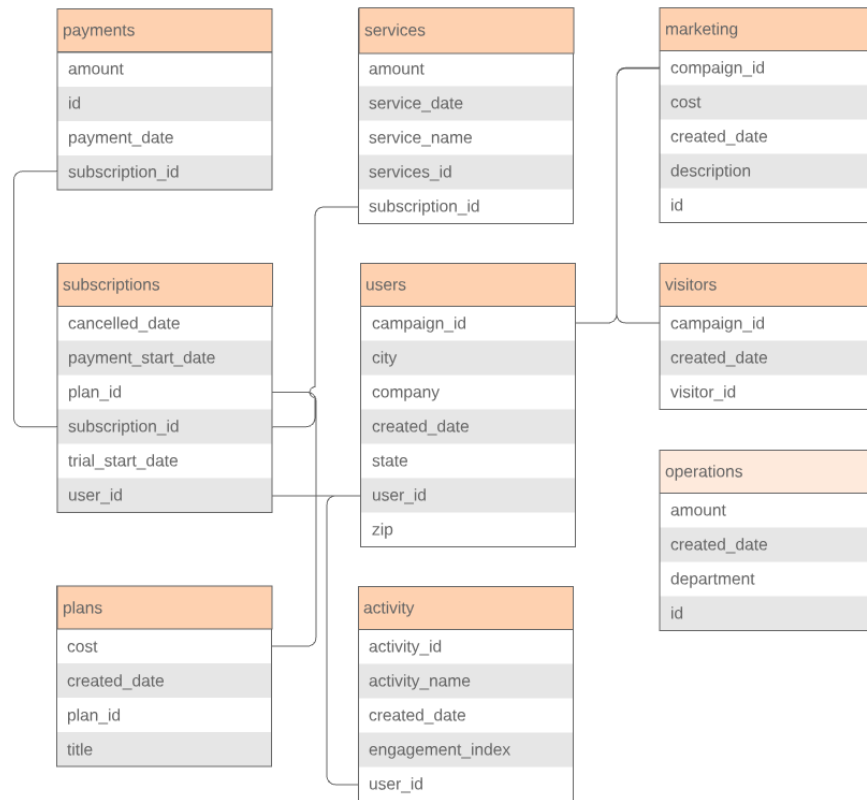
However the biggest challenge can be the “How”, as in how do we get the data to build the actual dashboard. The three data scenarios you will encounter are:

- When you have the data
- When you have messy data (most common)
- When you have no data

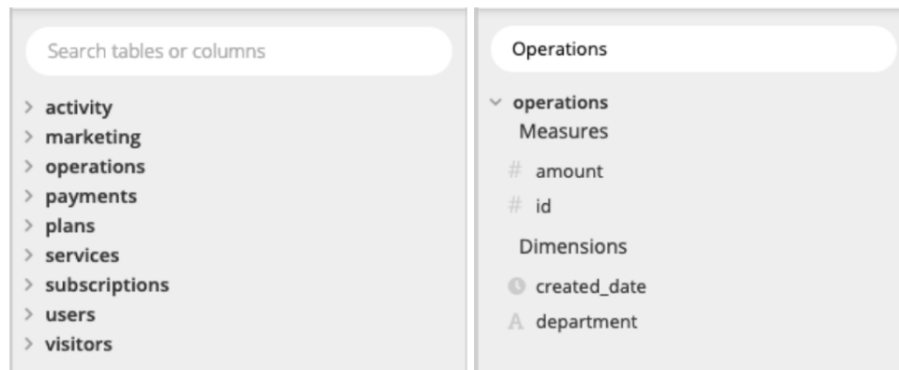
If you have the data this will be an easy step of the process. If the data is messy or it does not exist things will get much more challenging.

When you have the data

Now you have to find the data that you will use to calculate those metrics in your database. Review the schemas of the databases you have access to.



Finding where the data is that you need can be the hardest part of the process. Often times data is not well documented and what you need could be spread across multiple databases. First search for tables that have keywords from your Metrics Spreadsheet.



When you find something promising such as a table with the same keyword you searched or contains a field that matches your keyword do a quick query:

```
SELECT *
FROM operations
LIMIT 3;
```

id (integer)	department (text)	amount (integer)	created_date (datetime)
1451	Support	4525	2018-11-10 10:17:13.745322
1797	Payroll	4673	2019-02-01 16:18:49.132373
1776	Support	968	2019-01-28 04:15:21.862083

Once you get the results you can quickly see if it has the relevant information. In this case we can see it has a department column that looks appropriate, and the amount column may be the cost data we are looking for.

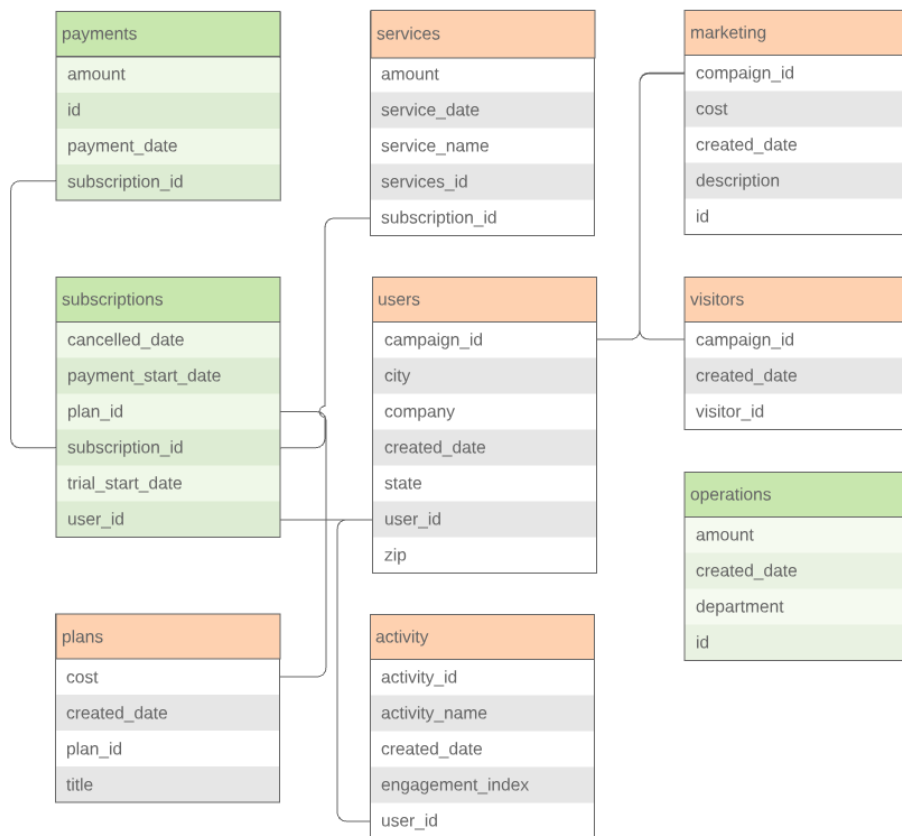
If a table such as this one looks relevant to one of the metrics write it down or put it directly in the Metrics Spreadsheet.

Your next step will be to determine what data you cannot find yourself. If you find a table but are unsure if it is the correct data put the name of the table in with a ? at the end. If you cannot find any relevant tables for a metric put three ? in the Metrics Spreadsheet.

Metric	Aggregation		Grouping		Filtering
	Formula	Datasource	Content	Datasource	
Operations Cost	SUM(Cost)	Operations	Total Department Month	Operations	Do not include marketing spend
Revenue	SUM(Revenue)	???	Total Month	???	
Subscriptions	Count(Accounts)	subscriptions?	Total Month	subscriptions?	Do not include free trials

If you have any questions about the tables or columns you have found, it is time to consult the Data Gatekeeper. I'd highly recommend coming to the Data Gatekeeper with the list of tables and fields you have questions about, using the metric spreadsheet is a convenient way to structure the conversation.

Go through each metric with the Data Gatekeeper, and explain what tables and fields you think you should be using and which ones you have questions about.



The Data Gatekeeper will confirm which tables and fields you have selected so far or will help locate the tables and fields you need. Some of the data you need might not be accessible to you due to access permissions. The Data Gatekeeper may grant you access or will provide feedback about how to work around this limitation. After locating the relevant tables, update the Metrics Spreadsheet.

Metric	Aggregation		Grouping		Filtering
	Formula	Datasource	Content	Datasource	
Operations Cost	SUM(Cost)	Operations	Total Department Month	Operations	Do not include marketing spend
Revenue	SUM(Revenu)	Payments	Total Month	Payments	
Subscriptions	Count(Accounts)	Subscriptions	Total Month	Subscriptions	Do not include free trials

We also need to specify the fields within the tables that will be used to make creating the SQL queries easy. Place the field names from the tables you found into the Formulas directly and put them in parentheses below your grouping categories in the Content column. Notice ‘Total’ does not have a column associated with it because it is the full aggregation.

Metric	Aggregation		Grouping		Filtering
	Formula	Datasource	Content	Datasource	
Operations Cost	SUM(amount)	Operations	Total Department (department) Month (created_date)	Operations	Do not include marketing spend (department)
Revenue	SUM(amount)	Payments	Total Month (created_date)	Payments	
Subscriptions	Count(subscription_id)	Subscriptions	Total Month (created_date)	Subscriptions	Do not include free trials (payment_start_date)

When you have messy data

Well you have data but it is messy. It may be obvious such as missing values. It may be mysterious, such as a value you have is different than what a user of your application is seeing. Let's explore what to do about different messy data problems.

Missing values

On any column that you will be using in a metric calculation you should check for Nulls and blank records. Here is an example query to get the Total number of nulls and nulls as a percent of total records.

```
SELECT COUNT(*)
FROM table
WHERE field is null
```

To check for blank values we can use:

```
SELECT COUNT(*)
FROM table
WHERE field = '' or field = ' '
```

You need to evaluate how to treat missing values.

- **Ignore:** leave the values as they are
 - This is the most common approach to dealing with missing data, you note the amount of missing values that exist and move on.
 - If there are high amounts of missing data in a field it is not recommended to use it in a calculation since it may no longer be representative of that field.
- **Delete:** remove any records with missing data

- Deletion is only recommended if the records are a very small minority of the data set so it does not skew the data.
- One legitimate reason for removing a record due to a missing value is if it has missing values in multiple fields.
- **Impute:** replace the missing value with a value
 - If the data is relatively normally distributed replacing the value with the average or median can be done if there are only a few records missing the value.
 - In some cases it is clear that leaving a field blank might indicate that the answer is 0, in these cases it is appropriate to impute with a 0

Regardless of which option you choose be sure to have the decision documented so others can reproduce the calculation and understand if they should take the data they are seeing to be 100% accurate or not.

Obviously Wrong Values

On any column that you will be using in a metric calculation you should also check for bizarre values. The Data School recommends doing a quick check on the highest and lowest values of any of the fields that will be used. You can do this using the ORDER BY clause.

```
SELECT *
FROM table
ORDER BY field DESC
```

```
SELECT *
FROM table
ORDER BY field ASC
```

This will quickly surface values that are way off if they are in the field.

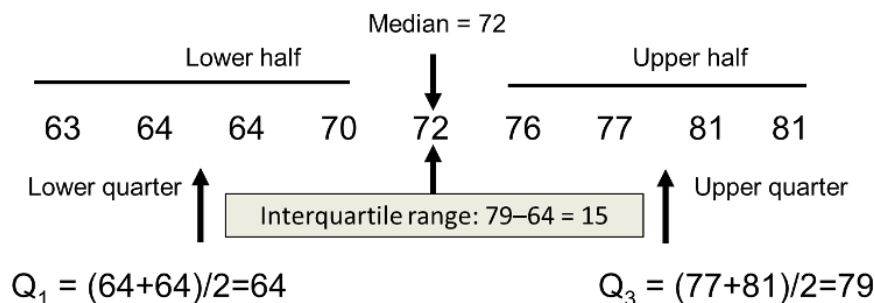
In text fields this is more difficult to detect, however there are some tricks here as well. They are more use case oriented tips.

- Phone Number
 - Common fake numbers 123-456-7890, 000-000-0000, 867-5309, and 999-999-9999
- Name
 - Common fake names: John Doe, Jane Doe,
 - Repetitive letters such as 'aa','bb'
- Birthday
 - Too old: pre 1919
 - Too young: 2014 and above

Potentially Wrong Values

While the previous method showed us obvious outliers there can be more subtle outliers that you may want to address.

Quartiles split a quantitative variable up into four equal sections. The interquartile range (IQR) is the difference between the upper quartile (Q3) and the lower quartile (Q1), which cover the central 50% of the data. If a value is far enough from the middle, we might consider that an outlier.



http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_summarizingdata/bs704_summarizingdata7.html

A commonly accepted definition of an outlier is $1.5 * IQR + Q3$ and $Q1 - 1.5 * IQR$. In essence, if a data point is 1.5 times the IQR larger than the upper quartile, or 1.5 times the IQR smaller than

the lower quartile, it is an outlier value.

Here is an example query applying this formula to find outliers using IQR.

```
WITH orderedlist
  AS (SELECT Row_number()
        OVER(
          ORDER BY amount)AS num,
        quantity
  FROM   table)
SELECT num,
       quantity
FROM   orderedlist
WHERE  num > Floor((SELECT Count(*) AS c
                    FROM   table) * 0.75 + ( Floor((SELECT Count(*) AS c
                                                    FROM   table) * 0.75)
              - Floor(
                (SELECT Count(*) AS c
                  FROM   table) * 0.25) ))
      OR num < Floor((SELECT Count(*) AS c
                    FROM   table) * 0.25 - ( Floor((SELECT Count(*) AS c
                                                    FROM   table) * 0.75)
              - Floor(
                (SELECT Count(*) AS c
                  FROM   table) * 0.25) ))
```

Again in text fields this is more difficult to detect. Watch out for the following:

- Incorrect Grouping
- Misspellings of the same thing such as 'Hamburger', 'Hamburdr', 'Hanburger'
- Various capitalization 'Hamburger', 'HamBurger', 'hamburger'

SQL will treat each of these variations as unique, you can find these by grouping by the column the text values are in and reviewing any groups with very few records in them. This is potentially a sign that they should have been incorporated into a larger group but weren't due to misspellings and capitalization inconsistencies.

When you have no data

Sometimes you do not have the data in your database to calculate a metric. This can be seen as a huge roadblock but there are a few ways forward. We need to first ask ourselves the following to know what we can do.

- Do we care about historical data?
- Is the metric actually trackable?
- How much data is needed for this metric to be valuable?

Instrumenting new data points

If historical data is of no concern consider working with the Data Gatekeeper to talk with engineering about starting to track what you want. This will delay the project from completion based on the engineering resources available to help.

Consider the costs

- People time
- Data Complexity
- Database costs

If the metric desired here is for a statistical test, talk with the Data Gatekeeper about how much data will be necessary to do the test properly. Factor in how much time it will take to get that data and let the Point Person know. Often times people will want to draw conclusions as quickly as possible, do not let your dashboard get used incorrectly like this. Set expectations up front when instrumenting new data points and perhaps even note on the dashboard when conclusions can be drawn.

Proxy Metrics

If we want historical data but we do not have the data for the metric we wanted we can use proxy metrics. Proxy Metrics are metrics that give us the same or similar information to what we wanted.

Example:

The Point Person wants to know what customers think of the product.

Desired Metric: NPS

- This has not been implemented yet and is reliant on people filling out a form so there might not be a lot of data to draw conclusions from.

Proxy Metric: Return Visitors

- We could look at how often customers use the product as a proxy metric, we equivocate regular usage with our customers liking the product.

While Proxy Metrics are not exact they can give us a good estimation. Finding good proxy metrics can be dangerous because they can change the focus of your company.

Consider a company who has to wait a long time to see if their product was valuable to the consumer such as an education company. They provide a lot of knowledge but if their ultimate goal is job placement they have to wait several months to know if their education was effective for the student.

How can they measure if they are doing a good job before their desired metric event happens?

You could introduce surveys to measure how confident students are in their job prospects, you could measure graduation rates, or you could measure the amount of inbound interest in your students from outside employers. Each one of these Proxy Metrics which you might be able to assess sooner than the “did they get a job” metric will lead you to focus on different activities. So be cautious with Proxy Metrics

Summary

- Finding where the data is located can be challenging
- Most likely you will need to involve the Data Gatekeeper in your search, be prepared with what data you are looking for before engaging with them
- Sometimes you will not be able to find data due to your level of access and will need to consult the Data Gatekeeper.
- Fill out your metric architecture to know which table and field will be necessary for each metric.
- Determine how messy your data is and clean it appropriately
- If you do not have the data for a metric work with the Data Gatekeeper to instrument new data points or use Proxy Metrics

Build the Metrics



In the previous chapters, we filled out a metric spreadsheet. We took a vague ask from a Point Person and turned it into a well-defined list of metrics, calculations, and data sources. We will now use the completed metric architecture to create various SQL queries.

The columns of the metric architecture map to a SQL query.

Metric	Aggregation		Grouping		Filtering
	Formula	Datasource	Content	Datasource	
Operations Cost	SUM(amount)	Operations	Total	Operations	Do not include marketing spend (department)
			Department (department)		
			Month (created_date)		
Revenue	SUM(amount)	Payments	Total	Payments	
			Month (created_date)		
Subscriptions	Count(subscription_id)	Subscriptions	Total	Subscriptions	Do not include free trials (payment_start_date)
			Month (created_date)		

SELECT

FROM

GROUP BY

JOIN

WHERE

(only if different)

Take a look at a couple of sample queries we could create from this spreadsheet for the Operations Cost metric:

Total Operations Cost

```
SELECT SUM(amount)
FROM Operations
WHERE department != 'marketing'
```

sum (integer)
2628498

Total Operations Cost by Department

(When we introduce a GROUP BY statement we must include any column there in the SELECT statement as well)

```
SELECT SUM(amount), department
FROM Operations
WHERE department != 'marketing'
GROUP BY department
```

sum (integer)	department (text)
1263323	Payroll
335873	HR
260256	IT
769046	Support

Total Operations Cost by Department by Month

```
SELECT SUM(amount), department, TO_CHAR(created_date, 'YYYY-MM') AS month
FROM Operations
WHERE department != 'marketing'
GROUP BY department, month
```

sum (integer)	department (text)	month (text)
97026	Support	2018-10
47106	HR	2018-11
56772	Support	2018-07
55622	Payroll	2018-05
15338	IT	2018-06

Previewing rows 1-44 of 44

One of the beauties of SQL is that it can do the logistical work of finding the columns in the data sources, and it can also compute mathematical equations. Most other methods require you to first access the unaggregated data via SQL and export the data into the tool so that you can create the calculations. Since SQL is tied to accessing the database when the underlying data changes, you can rerun the query and see the latest data. This is more efficient than exporting data into another tool.

SQL Resources

If you are struggling with understanding how Aggregations or Subqueries work check out:

- [How SQL Count Aggregation Works](#)
- [How SQL Subqueries Work](#)

If you are running into errors or are getting 0 rows returned check out:

- [Debugging SQL Syntax Errors](#)
- [Debugging SQL 0 Rows Returned](#)

Checking your Queries

Do not assume your query is perfect. You should check it by looking at other peoples' queries and/or by having the Data Gatekeeper review it.

Check other people's Queries

Depending on the BI tool that you are using you can see other people's SQL queries. This can be very insightful. You can take note of data sources they used that you were not aware of. You can also see if other people have complexity in their queries.

Complex Query example:

```
SELECT DATE_TRUNC('day', "Payments"."payment_date")::DATE AS "Day of Payment Date",
SUM("Payments"."amount") AS "MRR"
FROM "public"."payments" AS "Payments"
WHERE ("Payments"."payment_date"::DATE BETWEEN {CALENDAR_INTERVAL.START} AND {CALENDAR_
```

```
GROUP BY DATE_TRUNC('day', "Payments"."payment_date")::DATE
ORDER BY "Day of Payment Date" ASC
LIMIT 1000;
```

Complexity in a query typically suggests the data is nuanced, messy, or certain business logic needs to be adhered to. If you come across a complex query that is for the same or a similar metric as the one you are working on, try reaching out to the creator. You should try to understand what the extra parts are all about so you can incorporate what is relevant into your own query.

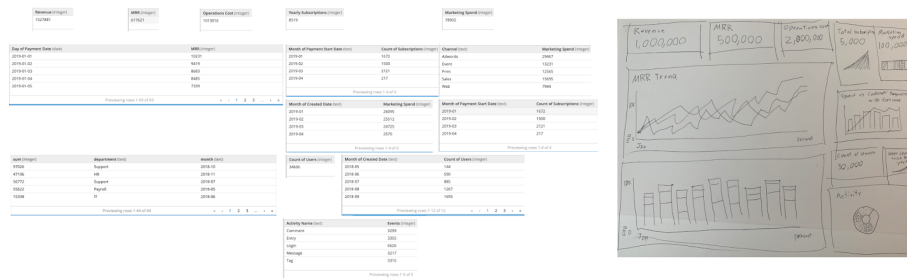
On the other hand, if other people have similar looking queries for similar metrics you are probably in the clear. However, you still will want to get someone else's eyes on it for verification.

Consult Data Gatekeeper

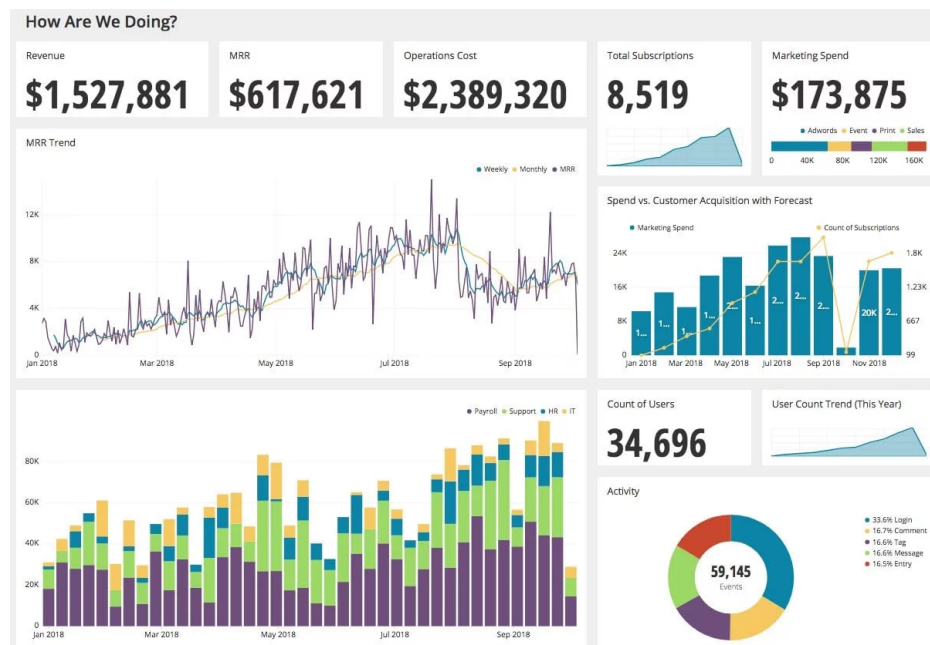
Getting a code review on your queries is a best practice. Reach back out to the Data Gatekeeper to validate your queries are calculating their metrics correctly. Having the metric spreadsheet facilitates this process since they can see your work and how you go to the query you wrote.

Build the Dashboard

Take the tables of data line them up with where they fit into your design.



Go through each table and create the corresponding data visualization in your BI tool. Put all the visualizations together into your final dashboard.

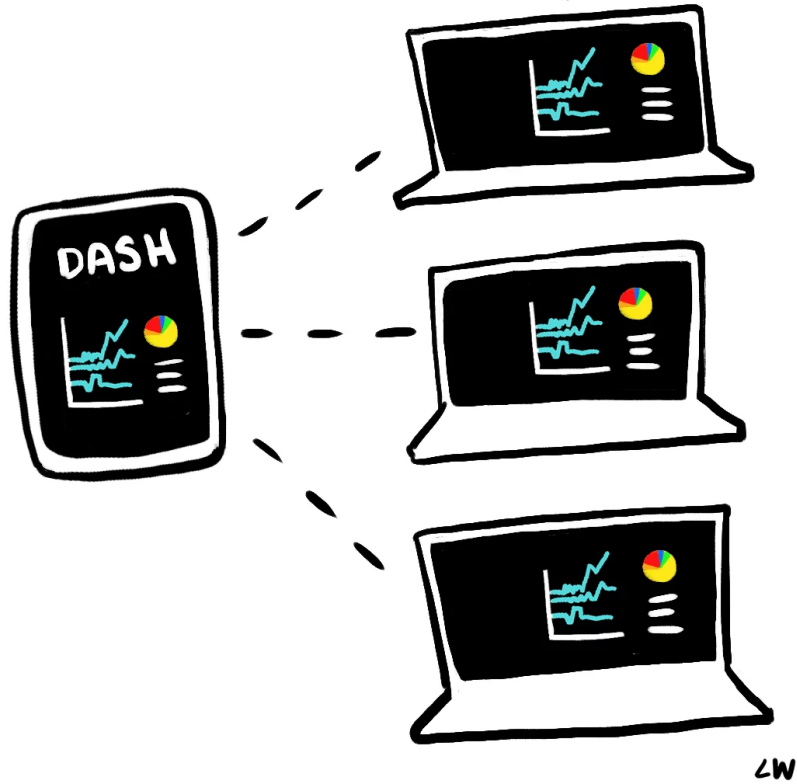


Summary

- Build metrics in SQL by plugging in the columns to their relevant part of a SQL statement
- SQL is required to get the data. Use it to calculate the metrics directly as well as to reflect any underlying changes
- Check your queries by evaluating other people's queries in your company and/or having the Data Gatekeeper review it

Deploy

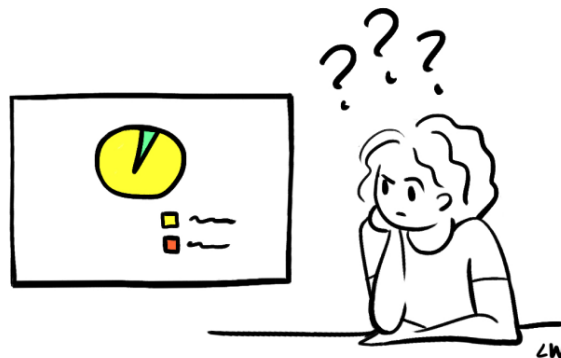
Sharing the Dashboard – Distribution Strategies



Now that you've built your dashboard you can begin to share it. Before you hit send take a moment to consider how to optimize the impact:

- Context
- Medium
- Scheduling

Context



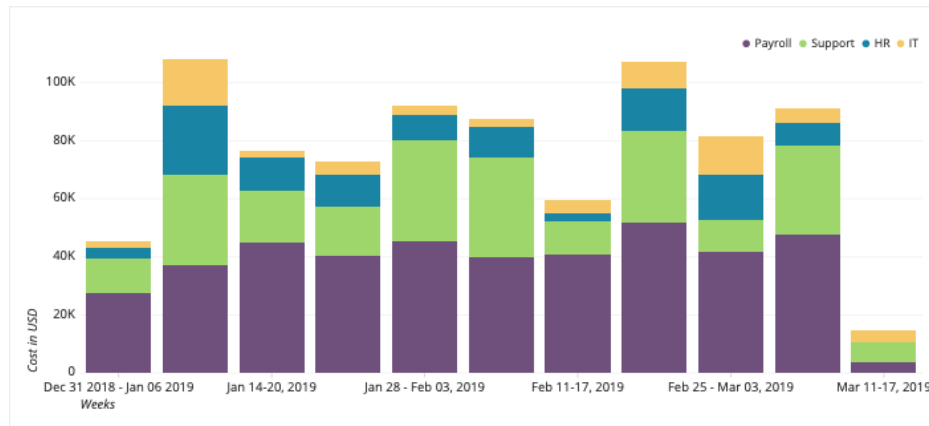
No matter how smart your audience is, you should design it to be readable by anyone in the company. The basics:

- Descriptive chart titles

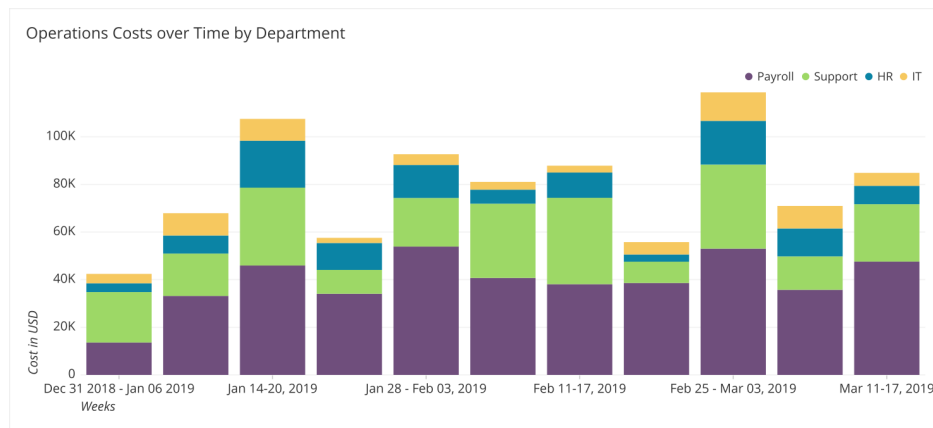
- Label axes
- Provide a key
- Add Definitions

Descriptive Chart Titles

Titles of charts should let the audience know what is being compared. Without a title it is unclear what they are looking at and it will take longer to process the information



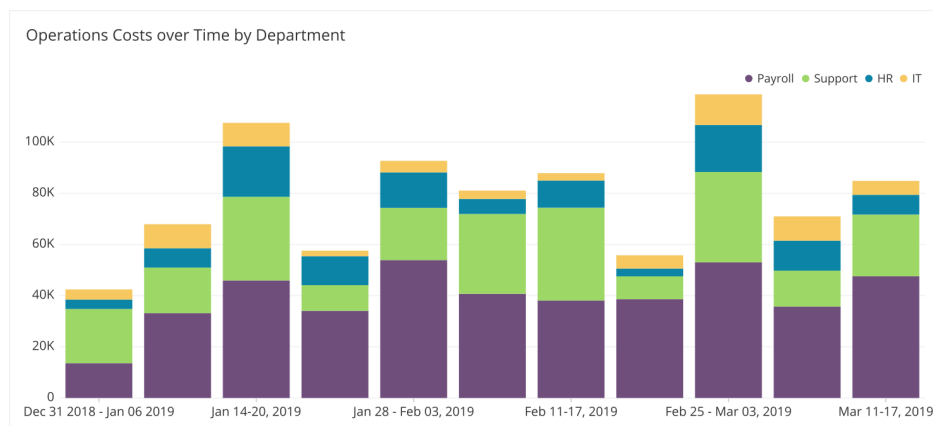
We can tell we are looking at cost in USD per week and we can see multiple departments from looking at the axes labels and key, however, the purpose is unclear.



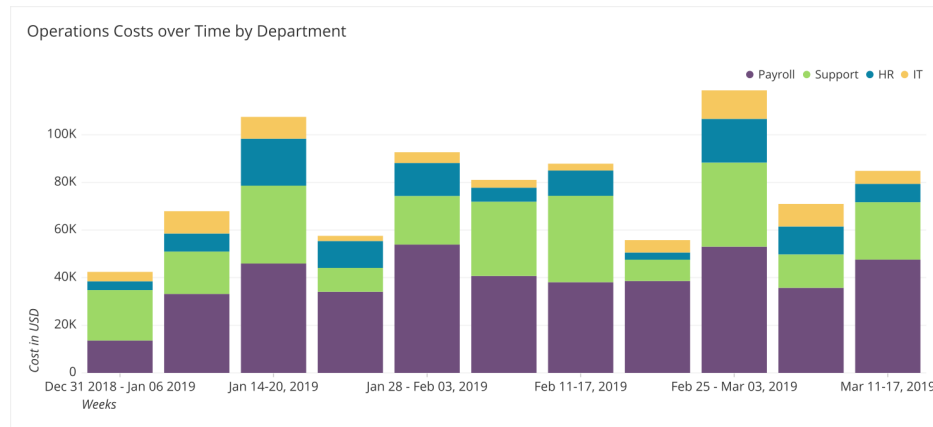
Once the title is added, the purpose of this graph is clear. We are evaluating the operations costs over time by department.

Label Axes

This is one of the most common pieces of context that is not included in a chart. Labels on axes should include a description and the unit for quantitative measures. If we look at the same graph as above without axes labels, we can see how it could become difficult to interpret.



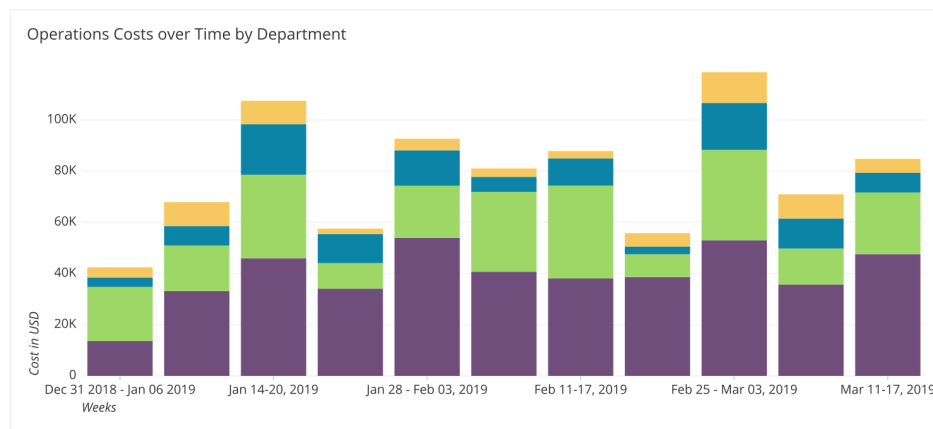
While the X-axis is fairly understandable given the labels, the Y-axis labels are a bit ambiguous. The title gives us a hint that we should be seeing costs but how are the costs measured? It could be different currencies, the number of employees, or the amount of hours worked.



Adding the Y-axis label makes it clear we are looking at costs in US dollars. While labeling axes can sometimes be left off, best practice is to include it first and then remove it based on feedback from the audience.

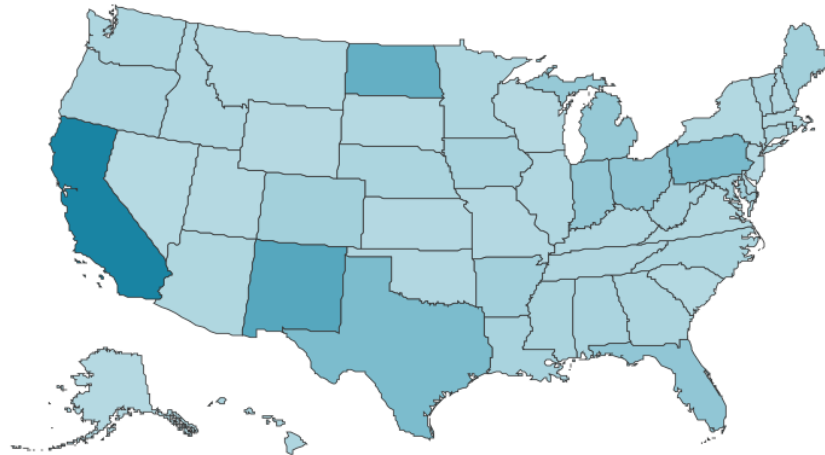
Provide a Key

Anytime color is used to delineate groups or represent a range of values, a key is necessary for determining what the colors represent. It is impossible to determine which bar belongs to which department without the key.



The lack of a color key is also very confusing when used to encode a quantitative range. As we can see in the chart below, we can infer that some states are different than others in terms of number of users. And we can also assume darker tones mean more than lighter tones but we do not know how much more.

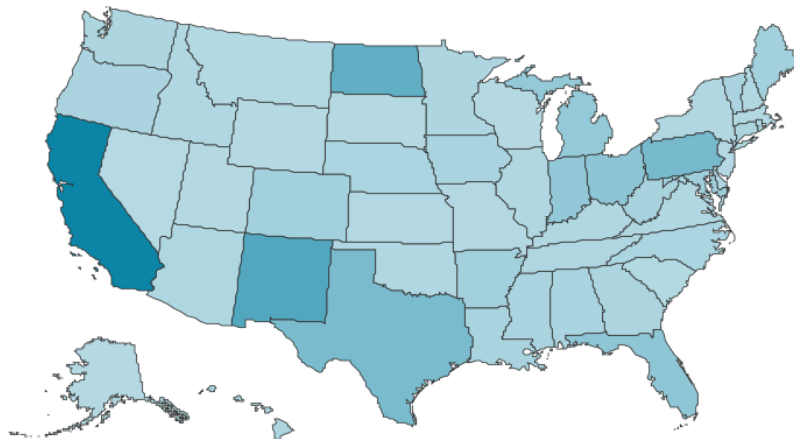
Number of users by Location



Once we add a key, the range of values becomes clear and the map is more useful for decision making..

Number of users by Location

1.41K 270K



Add Definitions

While some metrics are universal such as Total Revenue, some metrics such as MQL(Marketing Qualified Lead), DAU (Daily Active Users), and others are calculated very differently from company to company. The differences are typically due to excluding some groups from being factored into the metric. In such instances, providing notes in the dashboard or easily accessible from the dashboard is critical to interpreting the information accurately.

Net New Trials

946

Not measuring trials from existing customers or those that have not completed their password.

Adding a short description beneath the metric clarifies the information that is represented and its limitations.

Medium



Email

Emailing a dashboard is a straightforward way to sharing your work. However there are a few common mistakes you should avoid.

- Can this dashboard be shared externally (typically not)
- What level of access do people need to view the dashboard (typically they need more than you think)
- The more frequent the data is sent out the less likely it will be viewed (align the schedule with the goal of the dashboard)

You should state whatever policies are relevant in the email that includes the dashboard link. You should also set a schedule for emailing out the dashboard that is appropriate to it's goal.

Television or monitor

Dashboards can be stylized and formatted to be displayed on televisions or other large monitors throughout an office. You can see the problem with needing to scroll here, it isn't an

option. I have seen dashboards presented on multiple television sets mounted to the walls of bullpens, office spaces, conference rooms, and lunch/break rooms.



<https://www.geckoboard.com/learn/case-studies/dashlane/>

This type of distribution is casting a fairly wide net and consumers are going to have differing opinions regarding the dashboard. Sometimes those opinions are going to be in direct opposition of one another. In the final chapter, we will discuss options to remedy that problem. The Point Person responsible for the project is probably going to be your main contact regarding audience feedback. Be prepared to make significant changes on these type of dashboards once the dashboard is released.

From a technical perspective, there are a few other concerns.

- You will want to know how the television connects to the network so you can determine how to get it on the TV in the first place.
- You may need a dedicated computer or a wireless connection through Google Chromecast, Apple TV or similar technology.
- You will want to know the television's resolution as this will determine if you need to make any adjustments to your dashboard design to be more clear and readable.

Be prepared to test and troubleshoot these connections and displays before you officially launch the distribution.

URL from Application

Most software that helps you build Dashboard, support sharing via a URL. This software may be new to your audience. In which case, you will need to verify your audience has login credentials, some basic training, and access to the dashboard's URL so that they can view the dashboard. Many of these platforms provide the ability to set access permissions for every dashboard. Setting these permissions appropriately on a dashboard will prevent it from being accidentally deleted or edited by another employee.

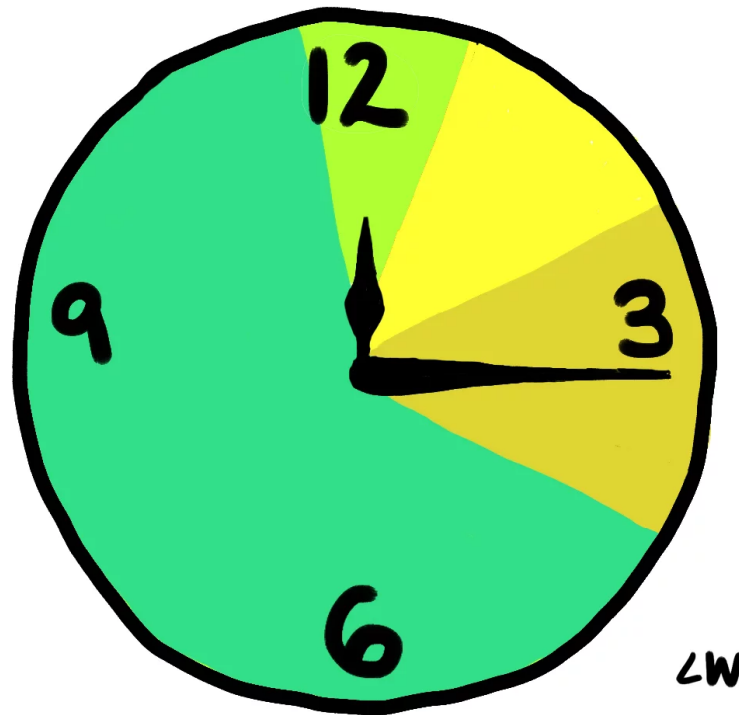
Paper copy

Occasionally an executive wants to receive a printed out version of the dashboard. While this can be slightly frustrating to accommodate, make sure the dashboard is delivered in the following ways.

- Timely
- Do not miss the scheduled delivery time. Executives often have a lot of meetings every day and this information might be critical to them making the best decisions.
- Securely
- You may not know the sensitivity of the information so after you've printed the dashboard put it in a folder and make sure the papers in the folder cannot be seen when the folder is closed.

- Personal Delivery
- At first, you will want to deliver the dashboard in person to keep up a necessary rapport for the feedback loop we will discuss later. Delivering it in person will always give you the opportunity to ask questions, even if the answer is short you will be able to gather some information on how the dashboard is fulfilling the executive's needs.

Scheduling



Scheduling should mimic the pace at which decisions need to be made. If the decisions that are made based on a dashboard are ad hoc then do not send out scheduled emails about the dashboard instead provide them links and let them check it whenever they need to. If people need to have context every morning of what is going on to start making decisions, then sending a daily email at 8:00am with the dashboard is appropriate.

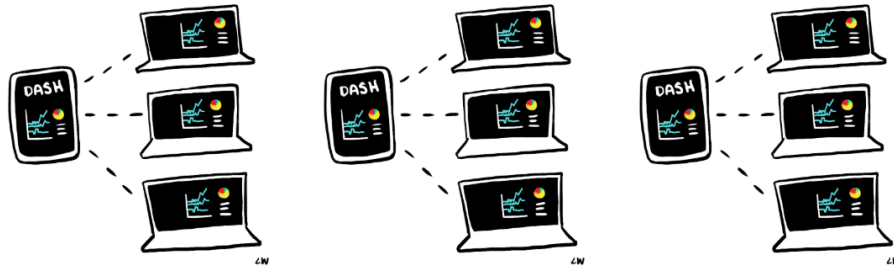
Finally, if people do not need to view the dashboard to make a decision unless a metric changes dramatically, set up an alert. Most dashboarding software supports being able to set a level on a chart that once it is crossed an alert will go out to an email of your choosing. Be careful with setting these, if you choose an artificially low number people could start receiving a lot of alert emails.

Sending people dashboards on a schedule that does not match their decision making needs could result in ignoring the dashboard and not using the data to support their decisions.

Summary

- Provide enough context for a dashboard so that your audience for clarity, accuracy and efficiency (can take it in with minimal questions)
- Choose a medium that aligns with the audience's expectations
- Use scheduling to help people make decisions not to remind them the dashboard is available.

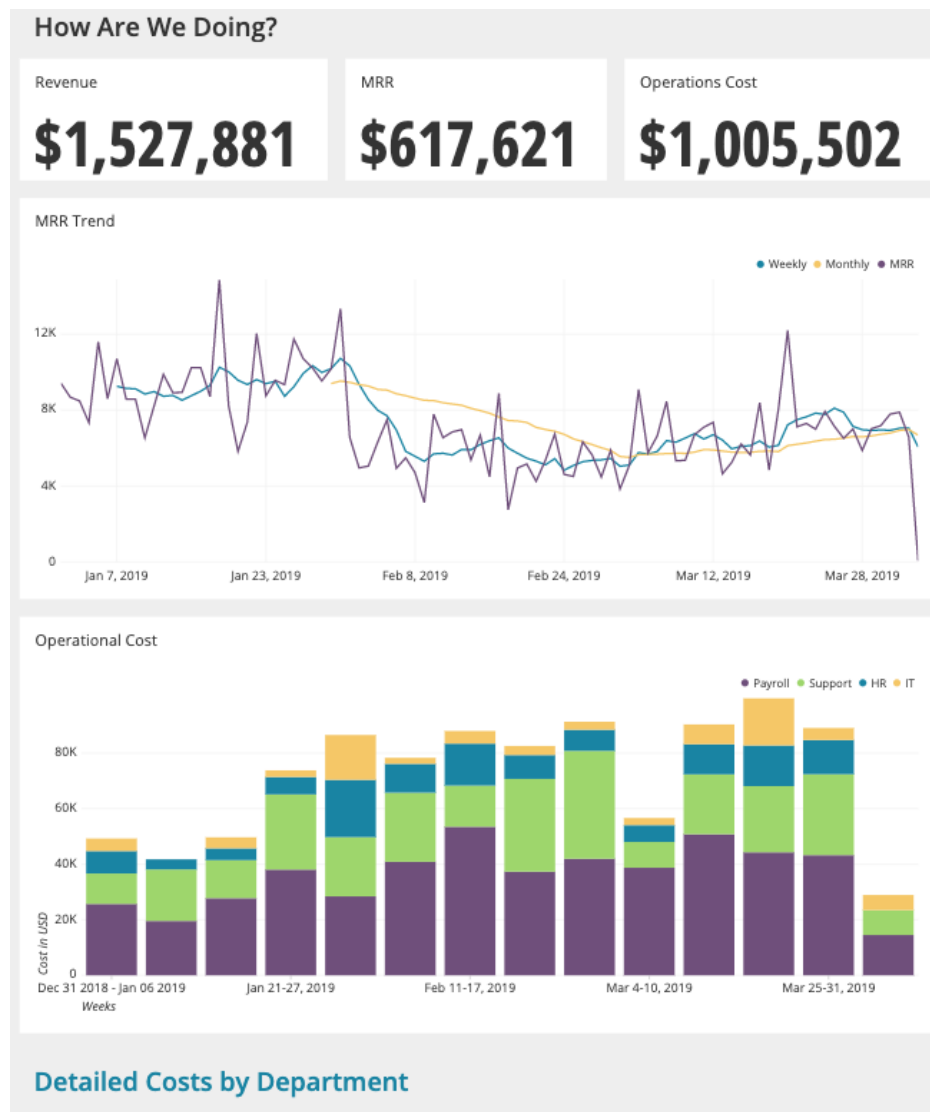
Scaling Dashboards



Once your dashboard is in front of its full audience how the dashboard is used is likely to evolve. This can be an expansion of decisions they would like to see supported or the number of groups who want to use it for their specific scenarios. To accommodate these changes there are some scaling strategies to consider using.

Linking out

If the feedback for the dashboard is to support more decisions consider if it is appropriate for the dashboard you current have or if you should start the dashboard design process over to create a second dashboard to support these new decisions. You can then provide links on the original dashboard to link to these new dashboards.

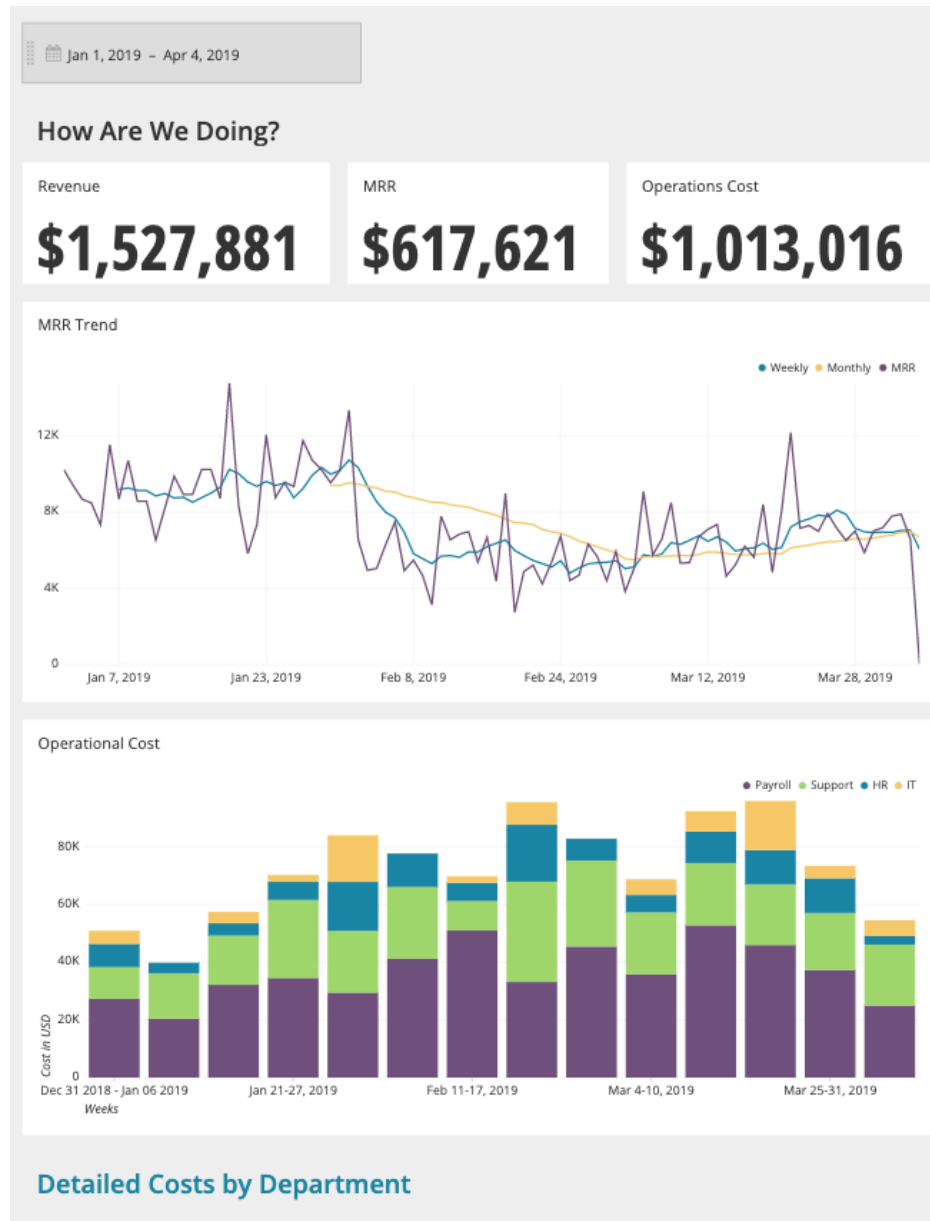


Here we have an example where at the bottom there is blue text that links out to another dashboard.

There are also technical benefits to linking out. Keeping the number of queries limited per dashboard will keep the dashboard loading quickly.

Interactivity

If the feedback for the dashboard is to support more groups' specific scenarios you will need to incorporate interactive features. This means having dropdowns for variables so that multiple situations can be evaluated using the same framework of the dashboard.



Here we have an example where at the top there is an interactive element to change the date range we are viewing.

When you introduce interactivity it is a best practice to turn off any auto refresh if you have multiple variables you can set. This will limit the amount of queries being performed until you confirm to refresh the dashboard.

Optimization

Regardless of how it evolves if the usage on a dashboard is high the demand on your database will likely increase. You need to ensure the dashboard still loads quickly and that the work placed on the database is mitigated. This can be accomplished by:

- Optimizing queries
- Setting the schedule
- Removing unused queries

Optimizing queries

At Chartio our rule of thumb is that if a query takes longer than 30 seconds there likely can be something done to optimize your query. If aggregations are taking a long time go to the Data Gatekeeper to discuss creating a pre-aggregated table that you can query from. This sort of data modeling can drastically improve query performance.

In addition leave any data manipulation (truncation, casting, etc) until after the aggregation. This means that you will aggregate the data first and then apply the transformations to the aggregated data.

```
SELECT SUM(num), category
FROM table
GROUP BY CAST(category AS VARCHAR)
```

```
SELECT SUM(num), CAST(category AS VARCHAR)
FROM table
GROUP BY category
```

This is much more efficient to making the change to every record before aggregation. You can check out more Query optimization strategies here: [Optimize your SQL Query](#).

Setting the schedule

Most business dashboards are not using real time data. Data is delivered to the data warehouse where you are querying from on a schedule in batches, which is setup by the Data Gatekeeper. Set expectations with the audience about how “live” the data is. You can also discuss with the Data Gatekeeper the speed at which people need to make decisions based on the data so they can adjust the schedule for when data is loaded into the data warehouse.

Removing unused dashboards

While this dashboard may be getting heavily used, it is likely some are not being used as much. It is a best practice to regularly archive dashboards that have not been viewed in over 90 days. This is because the more dashboards are querying the database, the more the database slows down for everyone. We do recommend sending out an email first alerting people which dashboards will be archived so they can respond and flag any that shouldn't be.

Documentation

Beyond the audience there is another consumer of your dashboard. This is future analysts who are trying to build their own dashboards. This could also be a future you. Do the future a favor and document your queries so that they can be understood easily and any quirks can be identified quickly.

A few settings will make your impact on the future much higher. Make sure that other people have access to the query and that they can access the data sources used. If these are protected sources, include the information about how to get access from the Data Gatekeeper

Summary

- Link out to support more decisions
- Add interactivity to accommodate more individuals/groups needing to make the same decisions
- Set a refresh rate that mimics the rate of decision making so that it is not refreshing more than necessary
- Document your work so that you and others can learn from your query decisions in the future.

Making Sure Your Dashboard Always Gets Better



Continuous Improvement is more than just a phrase or a buzzword, it is the key to a useful dashboard. You should not iterate for the sake of iteration. Iteration should be informed by feedback from consumers of the dashboard and the Point Person of the dashboard.

Get feedback after launch

Collecting timely feedback from the users and consumers of your dashboard will help you improve the dashboard. There are a few options for opening feedback channels that you may wish to investigate.

Feedback Channel

You will want to open some kind of easy, written feedback avenue, to give users a place to comment, offer suggestions, and ask questions. Providing your email or slack username allows for quick communication but can get overwhelming. A tool like Google forms captures feedback into a sheet that you can review as needed. For clear tracking, you can go with a tool like JIRA and have people submit tickets when they have feedback. As the amount of effort increases to send feedback, the less likely you will get feedback so make it easy.

Negative Feedback

Everyone knows that a person is most motivated to provide feedback when things are going wrong. This type of feedback is powerful, but not always constructive. It is helpful to focus on the source of their frustration. Is it a design choice, is it the data, is it not useful for them, or are they having a bad day? Consider their feedback and weigh it against the purpose of the dashboard and make a decision. The best thing you can do to encourage higher quality feedback is to let the person who gave you the feedback know what you decided to do with their feedback

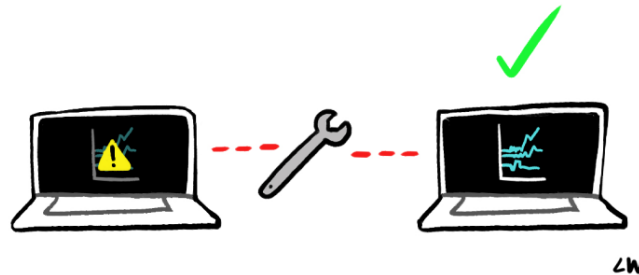
A common type of negative feedback is that something isn't working properly on the dashboard. This is a different type of feedback and it should be marked to grab your attention. Create naming standards for this type of issue such as: Place [BUG] or [BROKEN] at the beginning of your feedback so you can prioritize the fix. In some tools, you can accomplish this by tagging the feedback.

Audience to community



Moving your audience from a one way feedback channel to a community can unearth more valuable feedback. Use a community building tool like Slack, to get an insight into how viewers are talking about your dashboard with other viewers. These discussions will often provide more candid feedback. You can also solicit feedback on these channels. Once one person chimes in, others who are experiencing the same issues will tend to pipe up so you can judge the significance of the issue.

Iteration schedule



Set up an iteration schedule where you review the feedback you have received. A regular interval such as every month or every week works best. It is better to review feedback (other than bug/broken feedback items) on a schedule instead of ad hoc. This gives you the opportunity to prioritize your work and see if there are themes in the feedback.

Create a checklist, similar to the below, to review the dashboard in the scheduled review session.

Adoption and usefulness metrics of the dashboard

- Total dashboard views
- Repeat dashboard views
- Unique dashboard viewers

Accuracy check

- Do queries produce expected results?
- Have the underlying data sources or data models for this dashboard changed?

Data maintenance

- Check the load being placed on the dashboard.
- Optimize your SQL if necessary.
- Open and run the dashboard to see if any errors pop up
- Every so often your data sources may get re-architected for performance and system reasons. Make sure that the columns, tables, and databases still reflect the names that you provided when you wrote the initial queries.

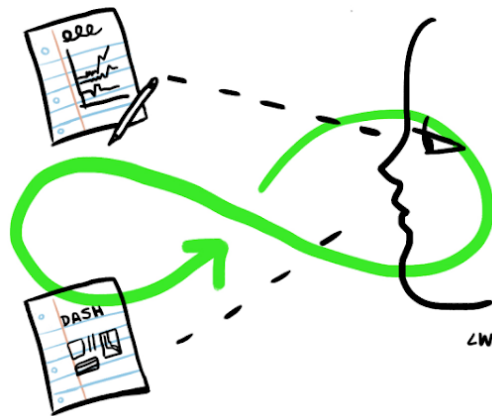
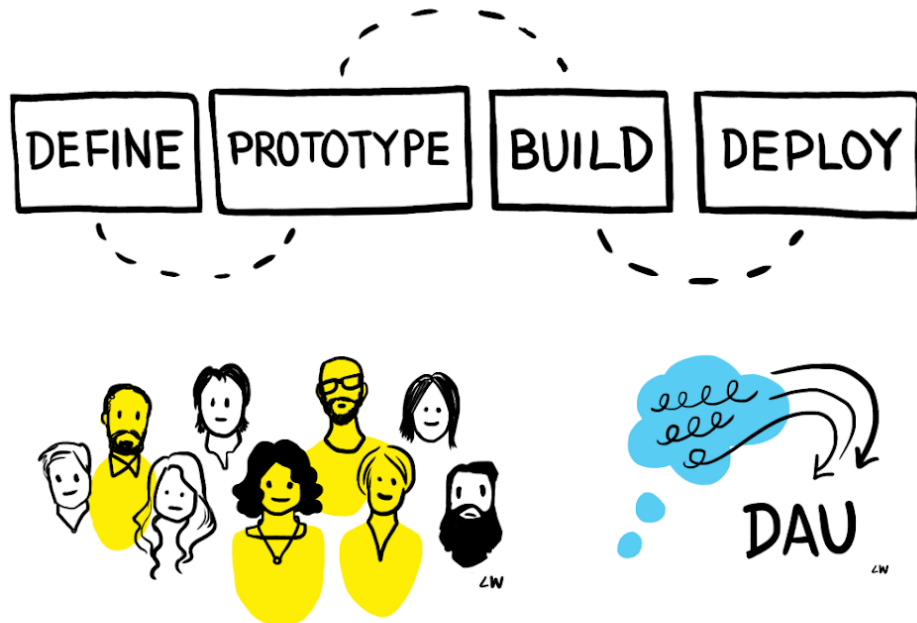
Summary

- Feedback from end consumer helps people make dashboards more useful and catch errors quickly
- Schedule time to review feedback to determine trends and prioritize work
- Regularly check out the health of the dashboard to see if people are using it, the data is still accurate, and if there are opportunities for optimization

Conclusion

Conclusion

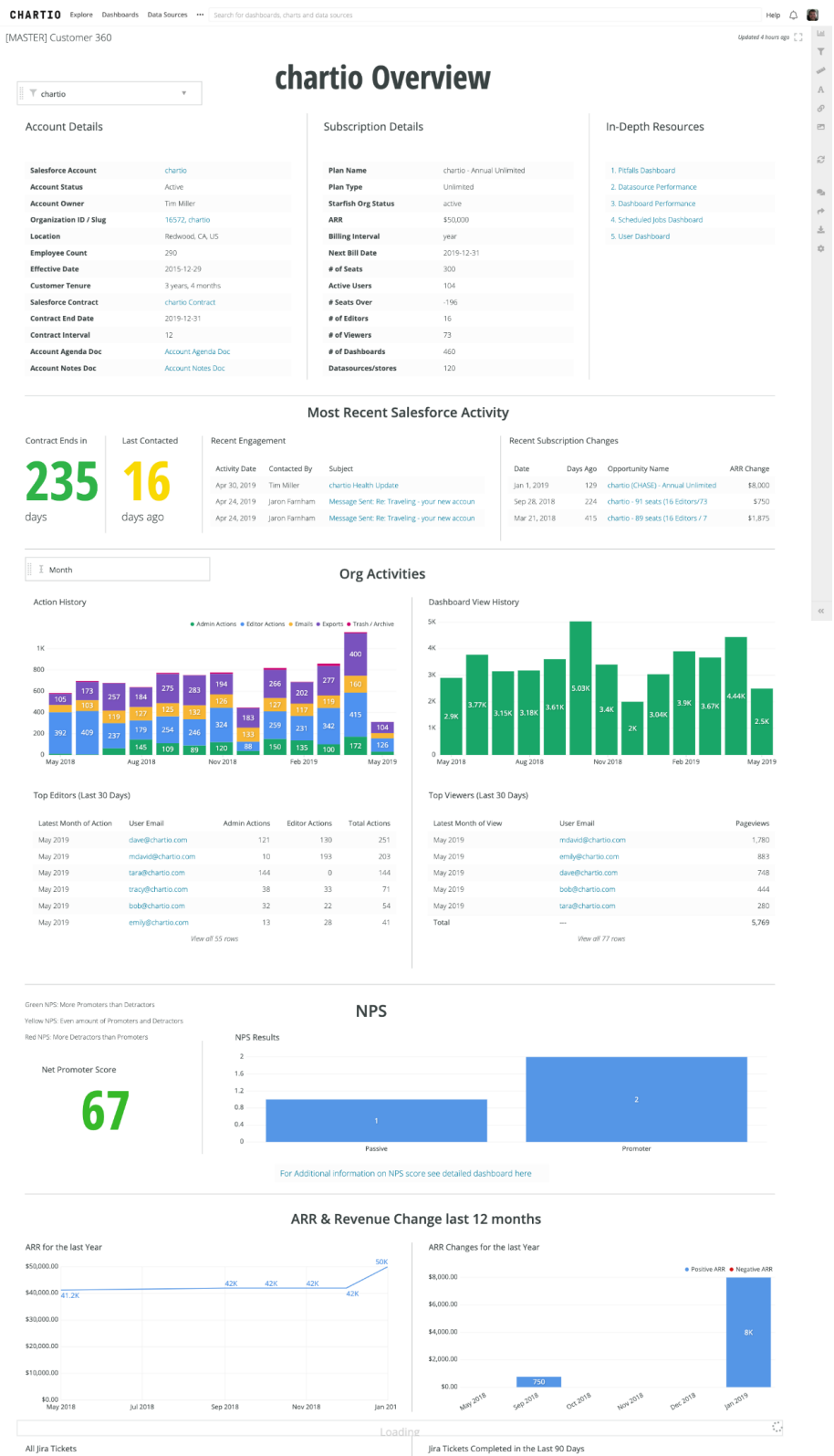
We have covered dashboard design best practices and outlined a process to create useful dashboards for any organization. You are now equipped to design and build a dashboards to help your organization be more data driven. We hope you enjoyed the book and look forward to hearing your thoughts in Slack.

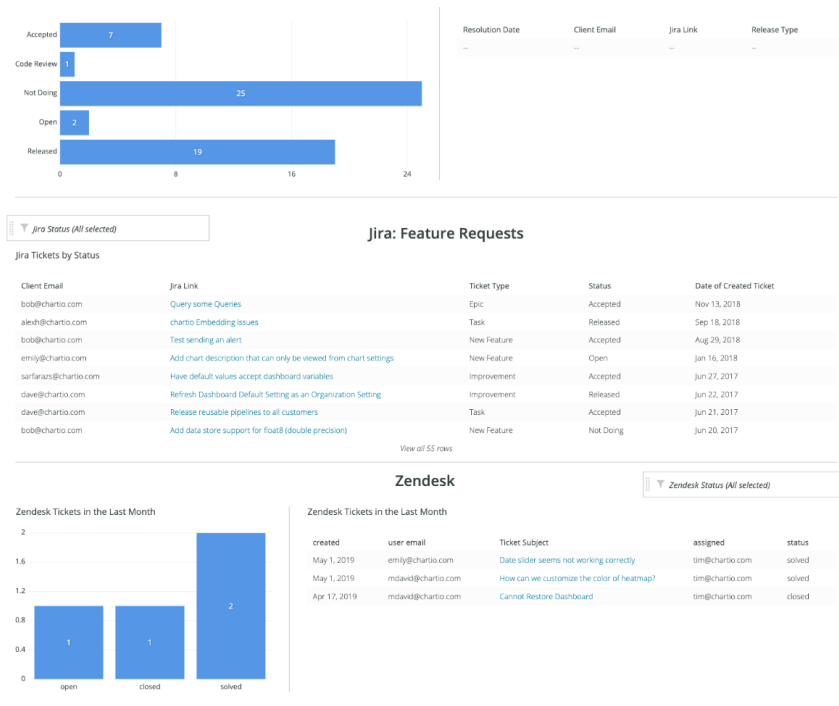


Design a Dashboard Example

A while ago I was working with my customer success team to organize a campaign to proactively uncover 15 common usage pitfalls customers can fall into so as to educate those who've fallen into it, or are near it, about the best practices.

Being very good at querying and visualizing data, they had created a dashboard with a ridiculous amount of charts on it. It was an only slightly organized view of almost everything under the sun about how a customer is using us.





Because it showed too much information in a disorganized way it was of little value (not shown in the image above is 14 more sets of visualizations about our customers you have to scroll to see).

Users of the dashboard came with a specific intent - to quickly determine what pitfalls a customer may be facing. This view required them to keep in memory what each of those pitfalls are, and be able to scan a bunch of information matrix style, and see if any of it might indicate one of those pitfalls.

These types of dashboards are what happens when you start by pulling data and think about the design and organization later. It's fun to explore and pull data, but when you do that it's too tempting to hoard it all on the same dashboard. Dashboards like this typically look like a lot of bar charts and tables on a screen. Properly designed dashboards on the other hand tend to be a lot more succinct and dense in their data display.

Ironically we'd ignored two of the very best practices we were building this dashboard to detect - keep shorter dashboards and for important dashboards design before you build.

I saw this as a perfect training opportunity, so we got everyone together and went over the practice and importance of dashboard design - using this very use case.

The Design Exercise

First we defined how the dashboard was going to be used. We decided that it was most needed when checking in on a customer, often before one of our monthly check-ins with them. We wanted a dashboard we could quickly scan and identify where the customer might be in or near a pitfall so we could discuss it with them.

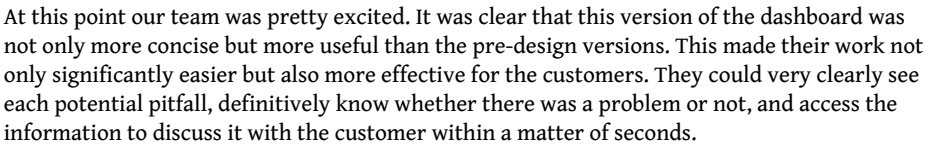
This made the decision of how to organize quite clear - the dashboard would be structured as a list of the pitfalls, with the info on that pitfall in each section. It had to be very quick to scan, and also have enough information on the issue to discuss it with the customer.

We then did something a little crazy for my team at the time - we put our computers away and we pulled out pen and paper. We divided up the pitfalls and each of us started drawing what we'd ideally have for each one.

For this example, we'll only focus on one of the pitfalls of customers not having Foreign Keys defined for some of their data source connections. It's very frequent that a database doesn't have relationships defined in its schema, and when that happens, Chartio's Visual SQL interface doesn't know how to do its magic and automatically join tables for you - and customers have a largely deprecated experience. (Note: As of writing this we now intelligently auto-detect foreign keys so this pitfall is no longer really possible!).

[illegible]

With these edits Tracy redrew the design and then built it out. It looked like this



2. Foreign Keys

Data Sources: Foreign Keys (Exds...

NOT SURE IF THIS IS GOOD OR BAD

ONLY TWO OPTIONS, A BULLET GRAPH WOULD BE BETTER

DON'T CARE ABOUT CSV'S

Data Sources: Foreign Keys

Data Source Foreign Keys by Dataset Count

THESE AREN'T IMPORTANT AS THEY HAVE NO QUERIES!

Datatype	Tables	Datasets	Foreign Keys
GoogleBigQuery	1000		0
GoogleBigQuery	1000		0
AmazonRedshift	656		82
AmazonRedshift	233	3251	59
Charlio Application	119	9912	159
PostgreSQL	118	1456	157

BAD DATA - NOT USEFUL

NO NEED

view all 159 rows and 8 columns

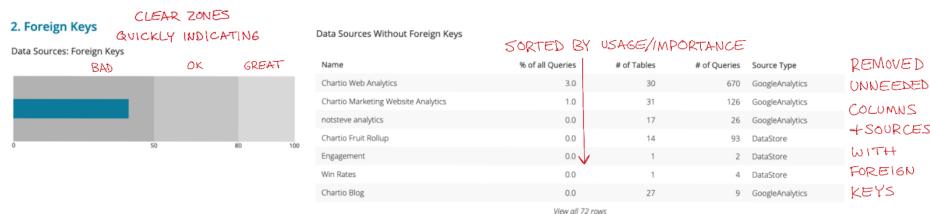
NO NEED TO SHOW SOURCES WITH FOREIGN KEYS

We determined again that the id column would never need to be known in this use case, and that the Creator Email was an unreliable and unnecessary column as who we really needed to speak to were the data governors of the organization not the original connector of the sources. The table on the right had expanded between design and implementation to now show not just

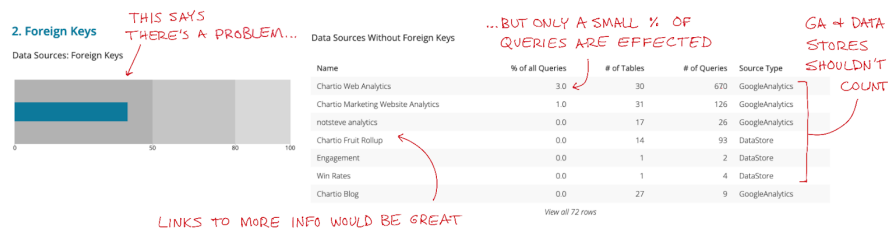
Data Sources without foreign keys, but those with foreign keys as well. The reason is that the builder thought it'd be cool to add a drill down feature where you could click the green or the red of either chart on the left to filter to the sources that respectively either had or didn't have foreign keys. It was cool, but it just wasn't too important.

The second chart had crept in with an attempt to not look at CSV files as a data source, but we weren't brave enough to go all the way to fully remove the data. Instead we moved it into it's own bar on a new chart.

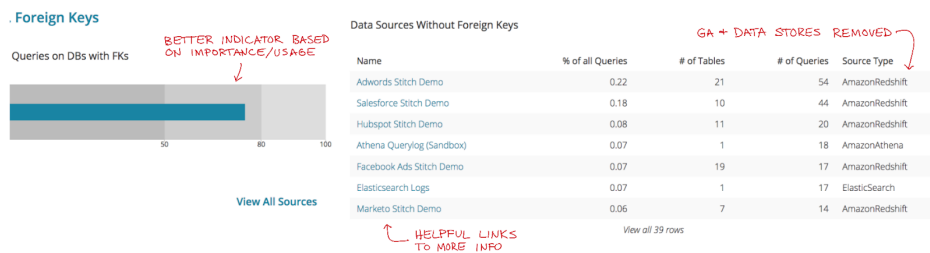
With the above changes we came up with a single bullet graph that would show what percent of relevant sources had foreign keys, and have clearly marked zones of where there's a significant issue.



We decided that the table should be sorted with those having the most queries first. We showed this as a % of all queries so you could quickly determine if it's a heavily or unused source, and therefore make a decision on whether it's worth talking about. The # of Tables and # of Queries were also considered valuable pieces of information for the customer conversation.



After a while of using this version, we realized that many of the data sources without foreign keys were hardly used. And that a better key metric for identifying if there's an issue was the percent of queries that are being run on a source without foreign keys. We also realized that GoogleAnalytics and our DataStores shouldn't count - as they're unique and never can have foreign keys.



So we changed the bullet graph, and added a few links to more information for each database, as we found that helpful in our customer conversations.

Phew! So many iterations for what ends up being just a bullet graph and a table. And this was for just one of the pitfalls, we had 14 others. The result has been hugely valuable though, as this is a critical and highly used view into the health of a customer.

As you can see in this example, to build such highly useful dashboards involves a lot of steps and skills. To build a well designed dashboards you need a good process and a number of key skills including visualization best practice, analytical design thinking and implementation/query knowledge.

This book was designed to well organize and educate on those skills. We hope it helps you create more simple and effective dashboards so that you can help inform your organization,

driving intelligent decisions and operations.