

Introduction to Data Visualization

A Communities of Opportunity Workshop
June 2020
Nikki Gurley, Monitoring and Evaluation, PATH
ngurley@path.org



My name is Nikki Gurley, I work with PATH, a local global health NGO

I work on our evaluation team, which supports a wide number of initiatives and projects across the organization

One of my particular areas of interest is the use of data – and in particular how to equip people with the tools to take control of their own data and use it to tell their own stories – making it more participatory, inclusive, and user driven. That’s why I’m excited about data visualization and how it can be used to make data more accessible and useable.

My co-facilitator today is Kris Johnson from King County Public Health

Land Recognition

We gather today on the ancestral land of the first people of Seattle, the Duwamish.

We honor with gratitude the land itself and the Duwamish Tribe, and all of the Native people who call King County home.

This acknowledgement is one small step against the erasure of our Native communities. In addition to the Duwamish, King County is home to the traditional lands of the Muckleshoot, Snoqualmie, and Tulalip tribes and other Coast Salish people. Today, King County is home to a vibrant and diverse community of indigenous people from across North America and beyond.

Resources:

<https://www.duwamishtribe.org/land-acknowledgement>

<https://www.realrentduwamish.org/land-acknowledgement.html>

<http://www.ala.org/aboutala/indigenous-tribes-seattle-and-washington>

Workshop Objective

Learn the basic principles of how to effectively visualize data using a variety of graphics from charts to maps.

Explore considerations for various ways to present data, how to ensure your data visualizations are clear and easy to understand, and examples and comparisons of visuals.

What we will do today: discuss a general framework for designing data visualizations, and some foundational elements or considerations for designing visualizations. There will be examples and opportunities for participation and feedback

What we will not do today: we are not going to talk about how to use specific software like Tableau, or specific how-to's in Excel, though we will link to resources you can use for those

Agenda

- 1 Why visualize your data?
- 2 A framework for visualization
- 3 Purpose
- 4 Content
- 5 Structure
- 6 Formatting
- 7 Bringing it all together

Agenda

- 1 Why visualize your data?
- 2 A framework for visualization
- 3 Purpose
- 4 Content
- 5 Structure
- 6 Formatting
- 7 Bringing it all together

Data overload and biases



“The problem for me was not ignorance; it was preconceived ideas.” – Hans Rosling

There is a lot of data out in the world – ranging from things like traditional census data, to interviews, to photos

It can feel overwhelming to process – how do we make sense of it all? How do we make it manageable?

When people get overwhelmed, we often rely on our biases and pre-conceived notions. So even though we have the data, if it’s not accessible, or we feel overburdened, we ignore it.

A pioneer of data visualization in Global Health, Hans Rosling, would survey people on what they thought the state of global health – including so-called experts, donors, or career professionals. Even though the data was widely available, they often fell back on stereotypes or pre-conceived ideas that certain countries or certain populations were doing worse than they were.

Full talk from Hans Rosling:

https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen?language=en

So the challenge is not that the information doesn't exist: it's that when it's too overwhelming or difficult to process, we don't use it

What can we do with visualization instead?

Make the data more accessible and digestible!

Visualizations can:

- Tell a story
- Summarize large quantities of data
- Highlight trends and relationships in data
- Inform decision-making

This is where data visualization comes in. The goal of visualization then is to make data more accessible and digestible - and in doing so, redirect and take ownership of the narrative

Examples of how data visualizations can be used:

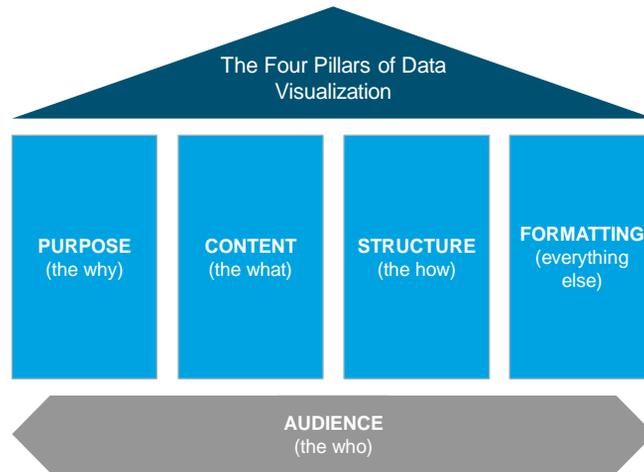
Elevate stories or voices that haven't been heard before

Inform what programs we should be offering communities, using data on community needs / desires

Agenda

- 1 Why visualize your data?
- 2 **A framework for visualization**
- 3 Purpose
- 4 Content
- 5 Structure
- 6 Formatting
- 7 Bringing it all together

Creating visualizations: the process



Adapted from the Four Pillars of Data Visualization by Noah Iliinsky

First, we use an approach to creating visuals called the four pillars of data visualization

At the bottom is our “audience” who is our foundation: they underpin everything we do. WHO you are creating your visualization for is an important consideration for the other four pillars:

Purpose, which is the WHY you are creating this data visualization.

Content, which is the WHAT- what data you are visualizing.

Structure, which is HOW you choose to represent your data, which charts you’re using

And Formatting, which includes everything else

The Four Pillars of Data Visualization

Purpose:

- Why am I creating this visual?

Content:

- What data am I using, and what matters?

Structure:

- How can we best present our data?

Formatting:

- What else is needed to make our graphic visually accessible?

Audience:

- *Who is our user and what do they need?*

Remember that your audience is still the foundation that underpins these other four pillars: Who are they, and what do they need?

Agenda

- 1 Why visualize your data?
- 2 A framework for visualization
- 3 Purpose**
- 4 Content
- 5 Structure
- 6 Formatting
- 7 Bringing it all together

Why are you making this visual?

Define your question

What is the question you are trying to answer with your visualization?

To help you identify your question, you can consider:

- What does my audience *need* to know?
- What would help my organization develop or improve our services?
- What is important and meaningful in my data?
- Is there a key question of interest already?

The first thing to determine is what is your purpose? What question are you trying to answer?

Maybe your audience needs to be made aware of the racial inequities in school suspensions

Or your organization needs to know what the community needs / wants are for an afterschool program

Sometimes you might already have identified a key question of interest based upon research you're already doing, or based upon a donor requirement

Again, the question is: what does your audience need to know?

Our visual will ultimately answer our question

For example:

Which school district has the fewest number of 10th grade students who met math standards in 2019?

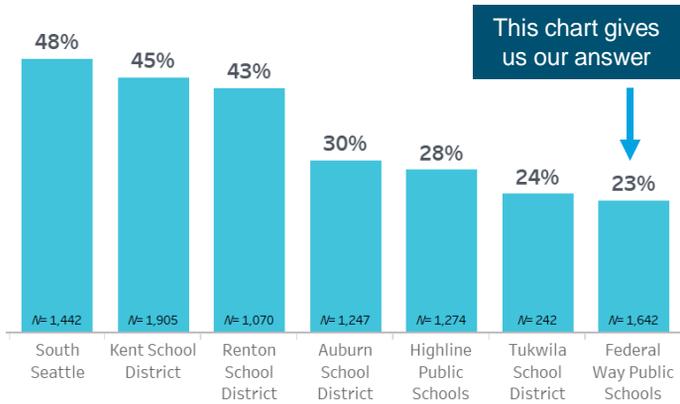


Chart from: <https://roadmapproject.org/data-dashboard/#3rd-11th-grade-assessments>

The reason it's important to identify our question is that the goal of our visual is to answer that question.

For instance, say the state has a bucket of funds for math improvement that they need to allocate. To ensure they're going to the places with the greatest need, we want to know which school districts has the fewest number of students who met math standards in 2019.

To answer the question, we build a visual that shows for each school district the number of students who are meeting math standards.

Our visual answers this question by showing clearly that on the far right, Federal Way has the fewest number of students meeting standards. Thus, we'd want to allocate funds there to help meet the needs of students.

This answers our question! The goal of your visual should be to answer your question.

Chart from: <https://roadmapproject.org/data-dashboard/#3rd-11th-grade-assessments>

Some possible purposes for a visual	Purpose	Question
	Compare two (or more) groups	What is the difference in the number of COVID-cases between Asian and Latinx populations?
	Show how things changed over time	How have student test scores changed since we implemented an after-school program?
	Understand the parts of the whole	Amongst our organization staff, how do people feel our board of directors is performing?
	Compare different geographic areas	What is the average annual income in different zip codes in King County?
	... and many more	

There are lots of possible purposes for a visual... there is no right answer here – it depends on what your question is and what your audience needs to know

We're going to cover four very common one's here and talk through these in the rest of our presentation:

- Comparing two or more groups
- Showing how things changed over time
- Understanding the parts of the whole
- Comparing different geographic areas

You'll notice there isn't a 'why' question here. The explanation of *why* something happened can sometimes be hard to capture in a visual, and this is where giving the appropriate context – whether in a presentation, or in text, is important for your audience to understand *why* the change occurred.

Agenda

- 1 Why visualize your data?
- 2 A framework for visualization
- 3 Purpose
- 4 **Content**
- 5 Structure
- 6 Formatting
- 7 Bringing it all together

What data am I using, and what matters?

Your content fundamentally is what data you're going to use. This could be a survey, it could be an interview, it could be from existing data, like the census, or any number of other data sources.

Communities Count website has a great data toolbox which has data sources that you can use: <https://www.communitiescount.org/data-toolbox>

Prioritize your content

Not all information is equally important

To help you identify your question, you can consider:

- What data matter?
- What relationships matter?

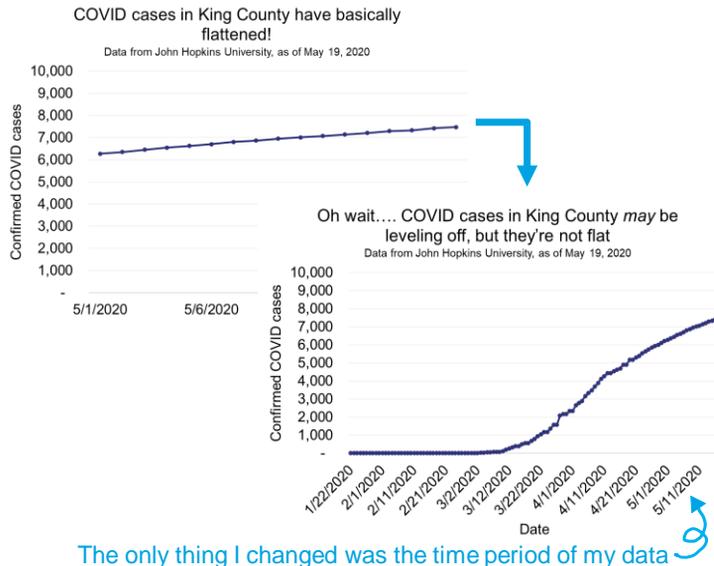
Avoid the kitchen-sink approach!

The thing we want to first thing about though is, what data is important in answering our question?

We talked at the beginning about how with so much data available we can end up overwhelmed.

When we pick our content we want to be selective so that we don't let our viewers get overwhelmed again!

Be honest:
minimize
distortion



Adapted from *Principles of Visualization* by Edward Tufte

The flip side of that is that we want to be honest.

Many of us have heard the adage that statistics will say anything you want them to. We don't want to manipulate our viewers

For example, this graph on the left looks like COVID cases in King County are fairly flat – they're nearly leveling off!

But this graph on the left (which is the exact same data), shows a much different portrait with a much sharper increasing trend.

The only thing I changed is my time period.

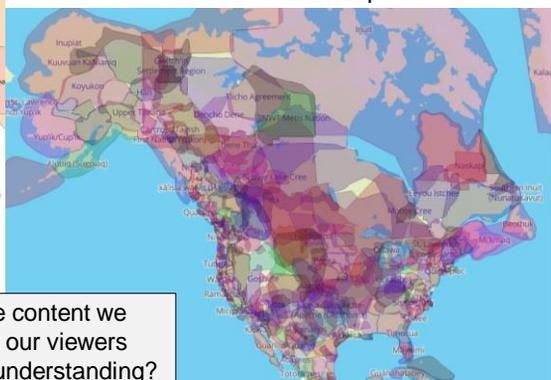
Be careful when you select your content that you are being honest about what you are displaying!

Choosing your content is important for equity: a case study

Most often this is the map of the US we see



But what if this was the map we saw?



How does the content we choose affect our viewers perception and understanding?

<https://native-land.ca/>

Thinking about our content is also important in considering equity.

This map on the left is the map most people have seen.

But what if we used this map on the right that shows indigenous territories instead?

It's the same data. But the content we're choosing tells our readers a specific (colonial) story or specific worldview.

It's important to be careful when you choose your content for this reason, and I encourage you to be reflective – think about:

What story is this upholding?

Does this reinforce harmful stereotypes or messages?

Is there missing historical context that's important?

Again, this is where you may need to pair further text or information to give your reader those additional insights and context of *why* this information is presented this way, or what world view is being told.

For a resource on indigenous territories: <https://native-land.ca/>

Agenda

- 1 Why visualize your data?
- 2 A framework for visualization
- 3 Purpose
- 4 Content
- 5 Structure**
- 6 Formatting
- 7 Bringing it all together

How can we best present our data?

We're going to spend much of the rest of our time talking about structure – how we decide what visuals to use, and which visuals are easier to interpret, and how to point people to what we want to see in a visual.

How many
'9's are in this
visual?

1	4	3	7	9	0	2	8	5	6
6	2	0	1	8	4	1	3	4	3
8	9	3	2	5	6	7	2	7	4
0	7	4	8	1	2	3	6	2	0
3	5	8	0	3	7	9	0	1	7
5	1	2	6	4	8	2	3	8	2
4	8	3	1	7	3	6	4	5	0
7	2	9	5	2	0	1	7	2	3
8	0	1	4	3	6	8	2	9	1
2	6	5	7	8	1	3	4	6	2

Consider, if you were to answer the question of how many 9's are here, does this visual give you the answer easily?

What about
now?

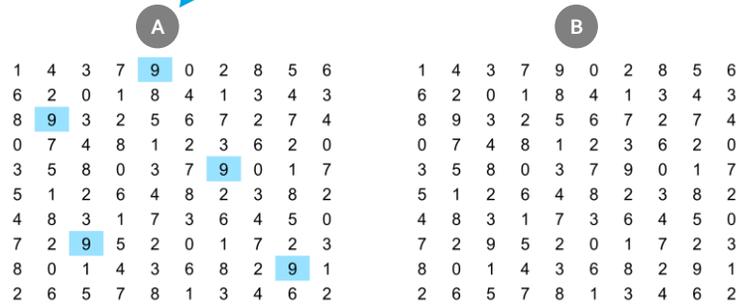
1	4	3	7	9	0	2	8	5	6
6	2	0	1	8	4	1	3	4	3
8	9	3	2	5	6	7	2	7	4
0	7	4	8	1	2	3	6	2	0
3	5	8	0	3	7	9	0	1	7
5	1	2	6	4	8	2	3	8	2
4	8	3	1	7	3	6	4	5	0
7	2	9	5	2	0	1	7	2	3
8	0	1	4	3	6	8	2	9	1
2	6	5	7	8	1	3	4	6	2

Consider, if you were to answer the question of how many 9's are here, does this visual give you the answer easily

The goal of our visuals should be to clearly communicate the answer to our question. The structure matters because it determines how easily people will be able to understand and identify that answer also

Which visual is easier to understand?

Why is this visual is easier to understand?



Because it takes advantage of how our brains process information

Adapted from 'Data Visualization: Cognition' presentation by Chris Adolph

Visual A answers our question much more easily

It's fundamentally the same content

But the structure of how we presented our data is different

The reason is that A takes advantage of how our brain processes information – it gives us a pattern to look at

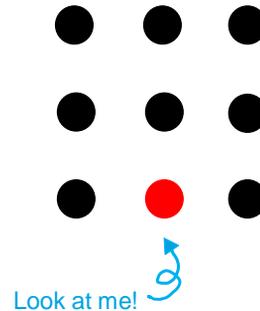
Why are some visuals easier to understand?

The brain can only absorb so much information at once.

We innately recognize patterns – through pre-attentive attributes.

Use **visual clues** to tell our mind “look here!”

For instance, color is a visual clue:



The brain can only absorb so much information at once visually

This is again the issue of information overload – when there is too much going on, our brains don't know what to do with it, or where to look. Think of the iSpy book effect.

However, our brains are very good at recognizing visual patterns through what are called pre-attentive attributes

Pre-attentive attributes are essentially visual clues that tell our mind to “look here”

For instance, color is a visual clue:

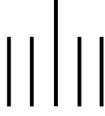
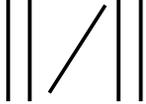
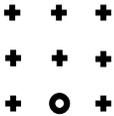
Between the black dots and the red dot, our eye is drawn to the red dot, because of the difference

The red dot doesn't fit the pattern!

This is a visual clue.

We want to use these visual clues to make our graphics easier to read – to tell the audience where to look, and how to answer the question

Examples of other visual clues we can use

<p>Length</p> 	<p>Orientation</p> 	<p>Position</p> 
<p>Size</p> 	<p>Shape</p> 	<p>And more....</p>

http://www.perceptualedge.com/articles/ie/visual_perception.pdf

Here are other common visual clues we use in presentations

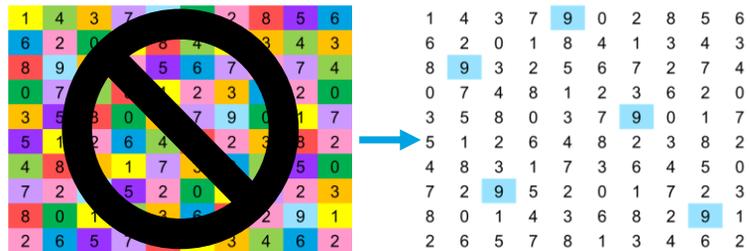
More visual clues can be found here:

http://www.perceptualedge.com/articles/ie/visual_perception.pdf

Don't overuse
visual clues

However, there's only so much the mind can process.

As you add each visual clue, the effect of each additional one diminishes.



It's important to keep in mind that there are only so many visual clues the mind can process

If we color code every number in our grid.... We end up back in the overwhelmed situation, where there's too much happening

We want to avoid using too many visual clues; focus on highlighting the important things

Focus on highlighting the answer to your question, your purpose!

Choosing chart structure to maximize understanding

We want to use chart types that take advantage of these visual clues!

Thus, different questions are (in general) better answered by certain chart types

The reason we want to talk about visual clues and pattern recognition is because this helps us determine what the structure of our visuals should be

What chart type best emphasizes the patterns we want to see?

Depending on what type of question we have, or what our purpose is, different visual clues are better than others at revealing patterns

So different questions generally are better answered by certain chart types

A short reference based on the 'purposes' we identified and visual clues	Purpose	Chart types
	Compare two (or more) groups	<ul style="list-style-type: none"> • Bar charts • Dot plots • Heat maps
	Show how things changed over time	<ul style="list-style-type: none"> • Line graphs • Bar charts • Gantt bars
	Understand the parts of the whole	<ul style="list-style-type: none"> • Pie charts • Stacked bar charts • Tree maps
Compare different geographic areas	<ul style="list-style-type: none"> • Choropleth (filled map) • Tile map 	

If we go back to our four purposes that we identified at the beginning, here are a few chart types that are good for each, based upon the visual clues they use

5 Common Chart Types

- Bar Charts
- Line Graphs
- Pie Charts
- Choropleth Maps
- *Text Tables*

We're going to talk about these five visuals

Some of these may be new to you, some may not

All of these can be done easily in readily available software's which we'll discuss later

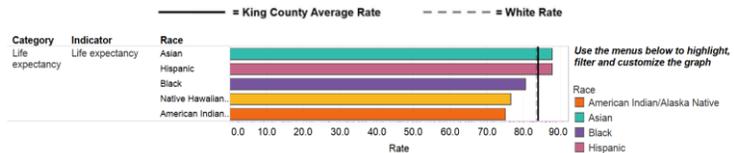
If you're feeling unsure or concerned about these different types of visuals, this is something the King County Public Health Team can provide support in also!

Bar Charts

Purpose: What is the life expectancy for communities of color in King County?

King County Health Disparities Dashboard Measure: (Rates)

About: This dashboard shows the rates of health and socioeconomic indicators for King County communities of color (bars) and compared to whites (dotted line) and the King County average (solid line). *Hover or click on the bars to see more information about each indicator.*



Note: The dotted line is the rate for White as the comparison group and the solid line is the rate for King County average as the comparison group.

- **Good for:** showing comparisons between groups, or showing rankings
- **Uses visual clues:** length and position

Chart from: <https://www.communitiescount.org/health-disparities-dashboard>

We're going to start with a bar chart

In this bar chart here, the goal is to show the difference in life expectancies among different racial groups in King County

You can see that each bar here represents one racial group to communicate the life expectancy for each group,

It uses length – how far the bar extends

It uses position – where is the end point of the bar

They've also added color, to differentiate between the different groups

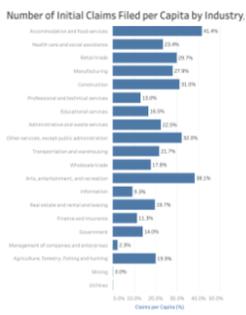
Bar charts are good for showing comparisons between groups, or showing rankings, because our mind can easily pull out the patterns of length and position to understand the data for each group

Here, you can easily see that American Indian's have a lower life expectancy than other groups

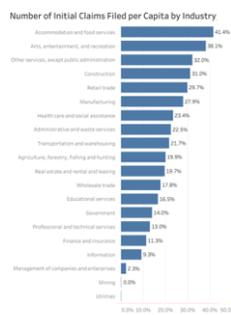
One thing you'll note is that this question doesn't ask or answer *why* are there disparities in life expectancy here? Again, this why question may be hard to answer in a visual, and would be benefited from a footnote or a text box which explains the structural factors which have led to disparate health outcomes

Chart from: <https://www.communitiescount.org/health-disparities-dashboard>

Bar Charts



Unsorted



Sorted!

Challenges	Tips and tricks
<ul style="list-style-type: none"> • Lots of bars make it hard to identify the takeaway • Really long bars if you have high values! 	<ul style="list-style-type: none"> • Sort order your data so the largest values are on top • Don't truncate bars! If needed, change baseline

Charts from: <https://www.kingcounty.gov/depts/health/covid-19/data/impacts/unemployment.aspx>

One challenge with bar charts is if you have a lot of bars it's not clear to know where to look or what the pattern is

For instance, in this chart of unemployment claims in King County by industry on the left – it's not immediately evident which are high or low

If we sort our bar charts so the largest values are on top (or by some other meaningful order!), this helps interpretation. You can see once sorted, it's a lot easier to read (on the right)

Charts from: <https://www.kingcounty.gov/depts/health/covid-19/data/impacts/unemployment.aspx>

Line Graphs

Purpose: How has traffic changed in King County once the Stay Home order was put in place?

Daily traffic percentage change from 2019, King County

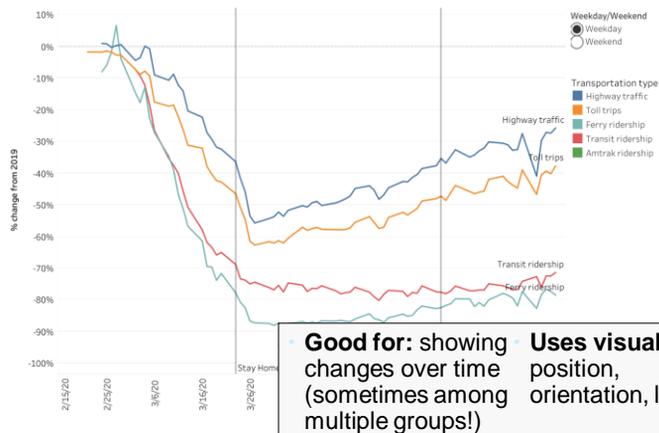


Chart from: <https://www.kingcounty.gov/depts/health/covid-19/data/impacts/traffic.aspx>

Now for a line graph. Line graphs are most commonly used to show a change over time – you’ll almost always see a date on the horizontal axis here

In this line graph, the goal is to show how traffic has changed in King County during the COVID outbreak. We’re using three visual clues here:

Position: where is the dot for each day plotted

Length: the line between each day

And most importantly, orientation: orientation is what tells us whether something is going up, or going down – this is part of what makes line charts so well suited to showing changes in time, is the inclusion of the orientation visual cue.

They’ve also added color, to differentiate between the different types of transit

These are particularly good at showing change over time – I frequently use these if I want to show how a program has influenced their constituents over time

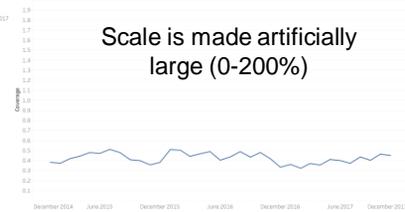
Here, we can see that movement declined sharply from the beginning of February to mid-March

Chart from: <https://www.kingcounty.gov/depts/health/covid-19/data/impacts/traffic.aspx>

Line Graphs



Scale is constrained artificially (30-50%)



Scale is made artificially large (0-200%)

Challenges

- Hard to identify the nuances, particularly if there are small changes

Tips and tricks

- Be careful to use a scale that shows the data in a meaningful (+honest!) way

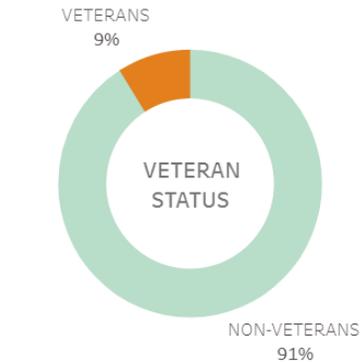
A key challenge with line graphs is the scale

This goes back to our question of content and making sure we're not being misleading:

On the left, my scale is artificially small and it looks like there's big jumps. On the right, my scale is artificially big and looks flat when it's not. We want to pick a scale that has meaning and is honest!

Pie Charts

Purpose: Among people accessing homelessness services in King County, how many are veterans?



• **Good for:** showing parts of the whole
• **Uses visual clues:** orientation and size

**technically, we consider this a donut chart – which is a pie chart, but with a hole in the middle!

Chart from: <http://allhomekc.org/data-overview/#who-is-accessing-services>

Now for a pie chart. This is one of the first chart types most people learn.

I've actually cheated here a bit – this is technically what's called a donut chart, which is a pie chart, but with a hole in the middle.

In this graphic, we want to know among the people accessing homelessness services in King County, how many are veterans? Pie charts are great for showing the parts of the whole, or what the distribution is *within* a group. To do that it uses two things:

Orientation: what's the angle here

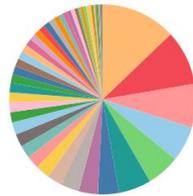
Size: How large is this slice of pie, compared to this slice?

One limitation is that it can be hard to tell exact numbers when you only use orientation and size – looking at this and I wouldn't know it was 9% if it wasn't labeled.

Otherwise though, it clearly answers our question by showing us that 9% of people accessing services are veterans

Chart from: <http://allhomekc.org/data-overview/#who-is-accessing-services>

Pie Charts



Too many!



Just right

Challenges	Tips and tricks
<ul style="list-style-type: none">• Too many parts makes it hard to identify all the slices!	<ul style="list-style-type: none">• Use pie charts only when your pie won't have too many slices

One challenge with pie graphs is we often are tempted to cut them into too many slices!

As we talked about before this brings us back to our data overwhelm – we can't process or find a pattern in all 30 of those slices, so they become meaningless

In general we want to stay with the fewer slices!

Choropleth Maps

Purpose: Which communities in King County have a greater proportion of people age 60+, and thus elevated risk for COVID?

Number of People age 60+ by City/Neighborhood

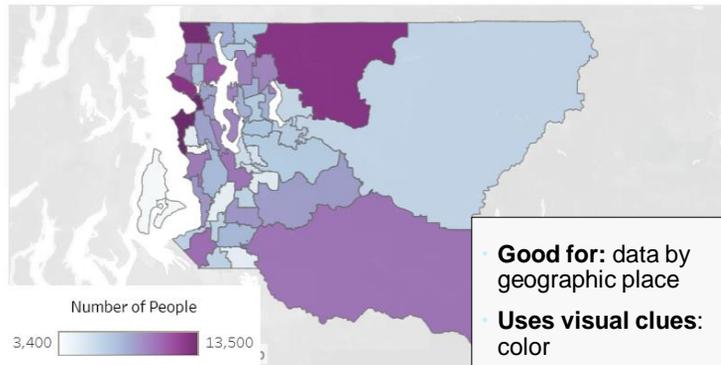


Chart from: <https://www.communitiescount.org/covid19vulnerable>

A choropleth is really just a fancy name for a shaded map. In this graphic, we want to know which communities have a greater proportion of people age 60+

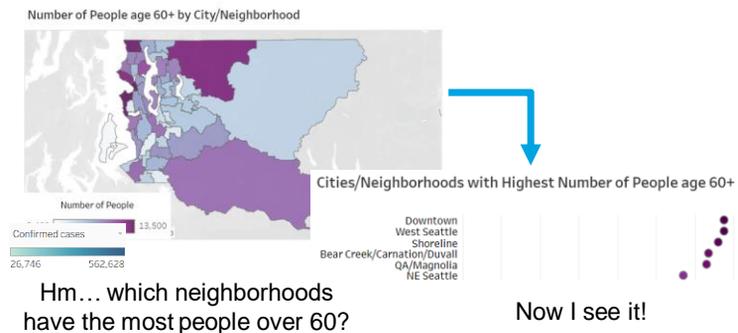
And you can see, the only visual cue we really are using in a map like this is color

We have our gradient that goes from light to dark and that tells us how many people

These are great for showing data by geographic place – where is something? What are the land boundaries?

Chart from: <https://www.communitiescount.org/covid19vulnerable>

Choropleth Maps



Challenges	Tips and tricks
<ul style="list-style-type: none">It can be hard to do specific comparisons using a map	<ul style="list-style-type: none">For precise comparisons between geographies, consider other chart types!

Chart from: <https://www.communitiescount.org/covid19vulnerable>

One key challenge is that it can be hard to compare things on a map.

For instance, if we wanted to know which neighborhoods have the most people over 60? Hard to say

If you want a specific comparison, you might want to think about using a different chart type – for instance that uses length or position like this dot plot here (the cousin of a bar chart) to show the specific numbers

It's important to think about what is the meaning of place when you're doing this. For instance, if we were looking at pollution, place is very meaningful as location and proximity industrial sites or similar would influence our findings

Chart from: <https://www.communitiescount.org/covid19vulnerable>

Text tables

Purpose: How many COVID cases and deaths are there in every municipality in the US?

Though it may not seem like it, these are also a visual!

Cumulative confirmed cases				Cumulative confirmed deaths			
METRO OR MICRO AREA	POPULATION	CASES	PER 1,000	METRO OR MICRO AREA	POPULATION	CASES	PER 1,000
1 Marion, Ohio	65,256	2,137	32.75	1 Albany, Ga.	153,009	152	0.99
2 New York City area	20.0 mil.	330,117	16.52	2 New York City area	20.0 mil.	19,211	0.96
3 Albany, Ga.	153,009	2,092	13.67	3 New Orleans	1.3 mil.	880	0.69
4 New Orleans	1.3 mil.	15,199	11.96	4 Fairfield County, Conn.	943,823	584	0.62
5 Fairfield County, Conn.	943,823	9,883	10.47	5 Detroit	4.3 mil.	2,371	0.55
6 Edwards, Colo.	54,993	488	8.87	6 Springfield, Mass.	631,761	285	0.45
7 Trenton-Princeton, N.J.	369,811	2,753	7.44	7 Hartford, Conn.	1.2 mil.	527	0.44
8 Grand Island, Neb.	85,088	629	7.39	8 Greenfield Town, Vt.			
9 Pine Bluff, Ark.	89,515	658	7.33	9 New Haven, Conn.			
10 Gallup, N.M.	72,290	519	7.18	10 Trenton-Princeton, N.J.			
11 Boston	4.9 mil.	33,039	6.78	11 Opelousas, La.			
12 New Haven, Conn.	857,620	5,811	6.78	12 Torrington, Conn.			
13 Sioux Falls, S.D.	265,653	1,671	6.29	13 Flint, Mich.			
14 Detroit	4.3 mil.	26,385	6.10	14 Boston			
15 East Stroudsburg, Pa.	169,507	1,015	5.99	15 East Stroudsburg, Pa.			

- **Good for:** showing exact values; not much else
- **Uses visual clues:** not many; position if sorted, or color if shaded.

Chart from: <https://www.nytimes.com/interactive/2020/04/23/upshot/five-ways-to-monitor-coronavirus-outbreak-us.html>

Lastly we have text tables! We often times don't think of these as visuals, because we use them so often in reports. But they are a visual.

They're really best at showing exact values; and not much else.

For instance in this chart where the goal of the NYT was to show every single case count in every single county, it can give us all those numbers

The challenge is they don't naturally use any of our visual clues. Two visual clues we can add to help interpret text tables:

Color – the areas with the highest case rates have darker colors

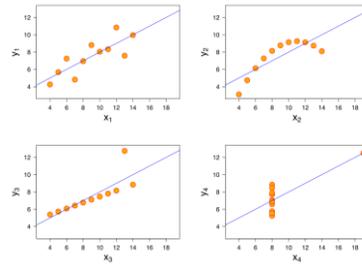
Position – we sorted our data, which adds an element of position – the highest rated numbers are on top here

Chart from: <https://www.nytimes.com/interactive/2020/04/23/upshot/five-ways-to-monitor-coronavirus-outbreak-us.html>

Text tables

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

This is the same data!



Challenges

- Relationships are not evident
- Lots of data points make it hard to identify what is important!

Tips and tricks

- Use a chart!
- Sort your data so it is ordered in a useful way
- Assign a color schema to your text table

One key challenge is that in a text table it's hard to understand relationships between two things

For instance, this data table on the left shows a bunch of values

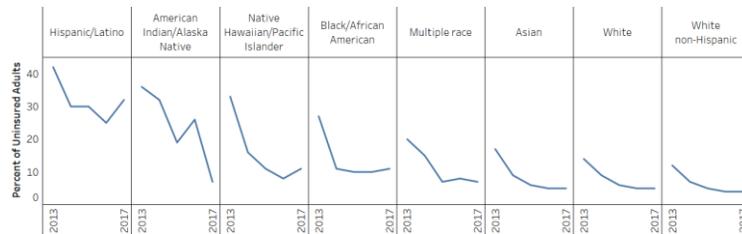
If we do a scatterplot on the right, we see a pattern emerge! But we can't see that in the table

If you know there's a pattern, try another chart type

Bonus: small multiples

Purpose: What has been the change in uninsured adults, by racial group, from 2013 to 2017?

Small multiples is using panels of the same graphic type repeatedly!
Any graphic type could be used (line graphs, bar charts...)



- **Good for:** showing large amounts of similar information
- **Uses visual clues:** depends on the graphic type you use. It requires the reader to only interpret one of the graphs!

Data from US Census Bureau, American Community Survey

The last bonus re isn't a chart as much as a strategy for multiple charts – what we call small multiples

This is taking panels of the same graphic type and using it side by side

It could be any graphic type – a bar chart, a line graph, or any other type

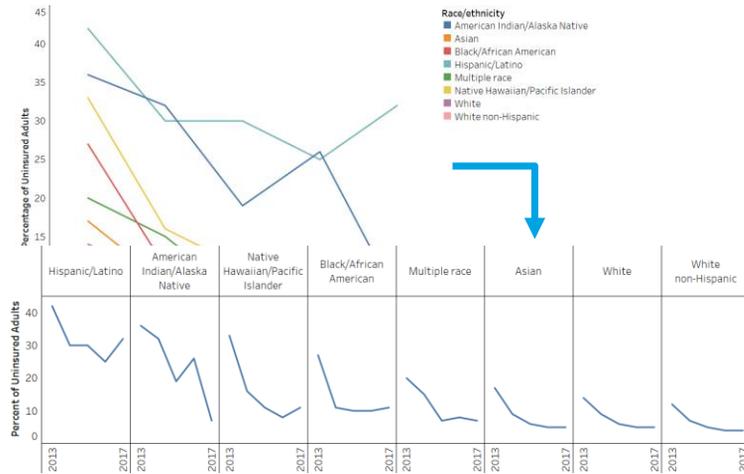
For instance, in this visual here we wanted to answer the question of what is the change in uninsured adults, by racial group from 2013-2017?

We gave each group their own miniature line graph and put them all side by side to see their individuals trends

The nice thing about small multiples is that it only requires the reader to interpret one of the graphs, and then they understand the pattern in each graph after

Bonus: small multiples

Many charts can be turned into a small multiples by adding panels....



Data from US Census Bureau, American Community Survey

Small multiples are really best for displaying large datasets where there is lots of similar information

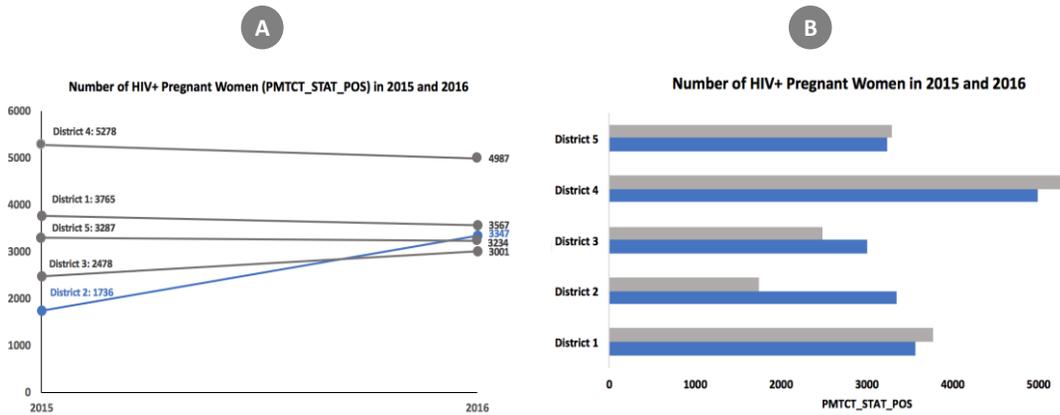
And you can turn an existing chart into a small multiple by just adding panels!

For instance, this graph on the left initially plots all the groups on the same visual. It's all the same information for each group though

To make it easier to see, we put them each in their own panel and put them side by side – small multiples! (graph on the right)

Purpose: which district has the biggest change in the number of HIV+ pregnant women between 2015 and 2016?

Which chart tells a clearer story? Why?



PATH's M&E PrEP presentation by Jenny Shannon and Jonathan Drummey

The goal of the visual here is to identify which district had the biggest change in the number of HIV+ pregnant women between 2015 and 2016

Which do you think is better?

There isn't a single right answer. Both choices are appropriate for visualizing how 2 or more numbers are alike or different. My preference if we are thinking about just identifying the district is Option A, because it highlights the one district with change. The orientation illustrates a single category increase.

But if we wanted to compare all of the districts, option B might be better because it shows bars for each district. The length helps us understand each district and compare them.

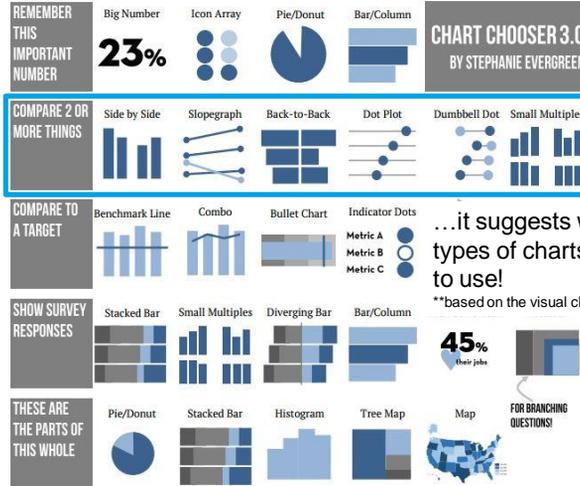
Again, the key question here is also who is your audience, their comfort, and what they need to understand. Some audiences may be more familiar with bar charts; others may intuitively understand a line graph.

What if I have
a different
question I
want to
visualize?

Many people have already done the work of researching and compiling different purposes (questions) and what chart types best use the visual clues to answer our questions....
Use a 'Chart Chooser'!

What is a Chart Chooser?

Based on our question or purpose...



...it suggests what types of charts are best to use!

**based on the visual clues they use

Stephanie Evergreen, <https://stephanievergreen.com/>

Some (free)
Chart
Choosers we
like

- [Data Viz Project](#)
- [Data Viz Catalogue](#)
- [Chart Guide](#)
- [iDashboards](#)
- [Depict Data Studio](#)

<https://datavizproject.com/>

<https://datavizcatalogue.com/>

<https://chart.guide/>

<https://www.idashboards.com/blog/2018/12/10/infographic-how-to-choose-your-charts/>

<https://depictdatastudio.com/charts/>

Agenda

- 1 Why visualize your data?
- 2 A framework for visualization
- 3 Purpose
- 4 Content
- 5 Structure
- 6 Formatting**
- 7 Bringing it all together

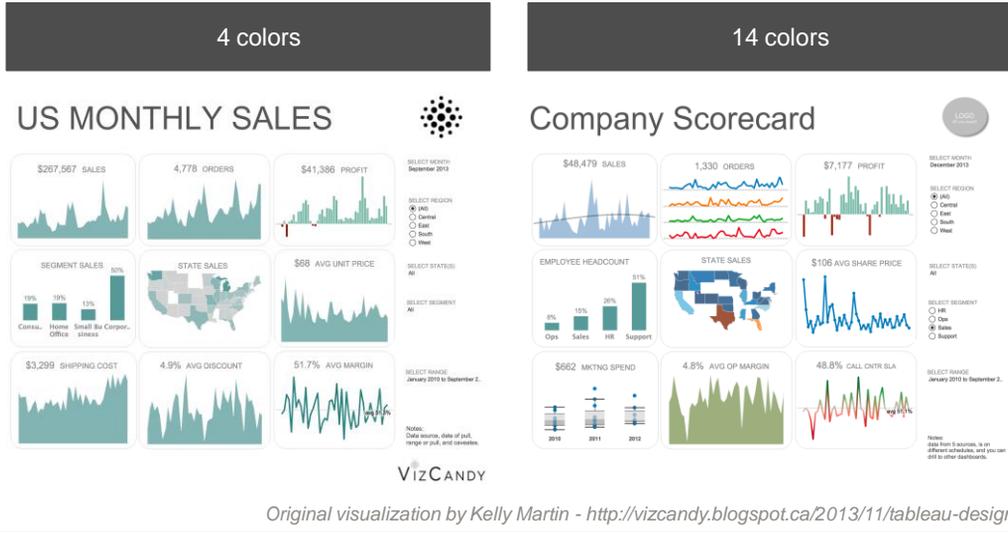
What else is needed to make our graphic visually accessible?

Color
Labeling
Font

Color: Some guidelines

1. Less is more. Stick to a few key colors.
2. Be intentional with the selection of your color scheme.
3. Make your color scheme accessible.
4. Choose a color scheme that supports different modes of delivery.

Color: Stick to a few key colors!



As before, we don't want people to be overwhelmed.

Color is one of our key visual clues, but if we have 17 colors, it no longer tells us where to look. Stick with a few key colors!

Original visualization by Kelly Martin - <http://vizcandy.blogspot.ca/2013/11/tableau-designs.html>

Color: Be intentional...

...and create meaning using branding (or other meaningful attributes)

A color scheme based upon the brands of beverage companies....

Beverage company A

Beverage company B

Instead of making assumptions about color, rely on choosing colors that are meaningful to the work you're doing, or the organization

Often times this means brand colors!

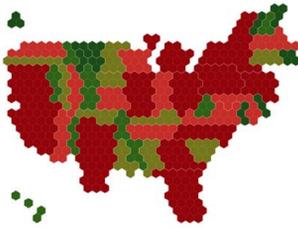
For instance, if I told you both of these colors were from a popular beverage company, you'd immediately know which companies I was talking about

Similarly, pick colors that will help your readers connect to your work – whether that's branding, or something your org uses

Color:
Choose an
accessible
color scheme

Make sure to select a color scheme that accounts
for color-blindness

Original visualization



Viewed by someone who is
red-green colorblind



Use the ['No Coffee' Chrome plug-in](#) to test your visuals

<https://chrome.google.com/webstore/detail/nocoffee/jjeeggmbnhckmqdhmgdckeigabjfbddl?hl=en-US>
Visuals from: <https://blog.datawrapper.de/colors/>

Another key element of choosing color is choosing a color scheme that is accessible to all

Most notably this means accounting for color blindness, which affects 5-10% of the US population

The most common form is red-green colorblindness, which means as we mentioned earlier a traffic light color scheme isn't a great fit

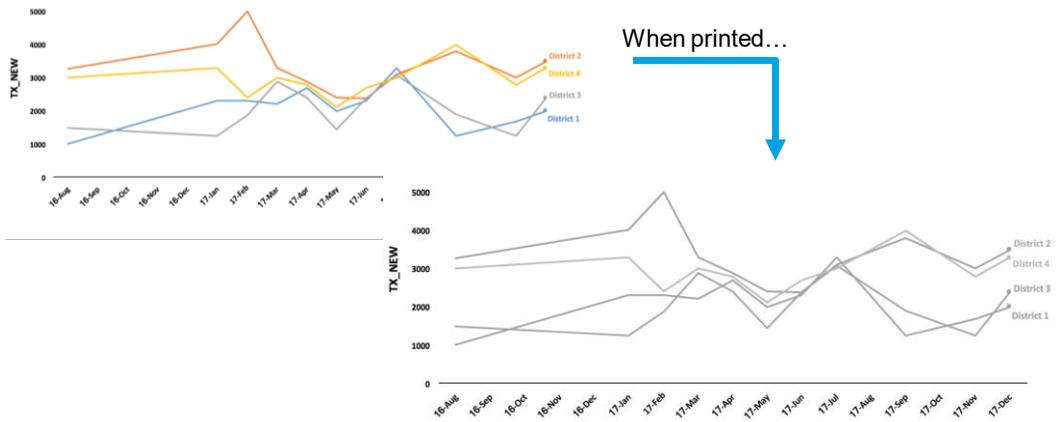
You can see on the left, we used the traffic light color scheme

But on the right is what it looks like for someone colorblind

You can use plug-ins which provide a simulator on Chrome to test your visuals

<https://chrome.google.com/webstore/detail/nocoffee/jjeeggmbnhckmqdhmgdckeigabjfbddl?hl=en-US>
Visuals from: <https://blog.datawrapper.de/colors/>

What if your visual is printed (in black & white)?
Make sure it's still readable!



PATH's M&E PrEP presentation by Jenny Shannon and Jonathan Drummey

The other thing you want to consider is how your visual will be presented

Frequently in my work, we'll print copies of reports or handouts for people to have – often times these are in Black and White

Many color schemes don't translate to black and white

For instance, this color scheme which used the Excel default (on the left) - looks like this when printed (on the right)

It's not longer possible to tell which lines are which, particularly in the middle where they crossover

So try printing your graphic to make sure the colors are BW compatible also

Labels and 'chart noise': Some guidelines

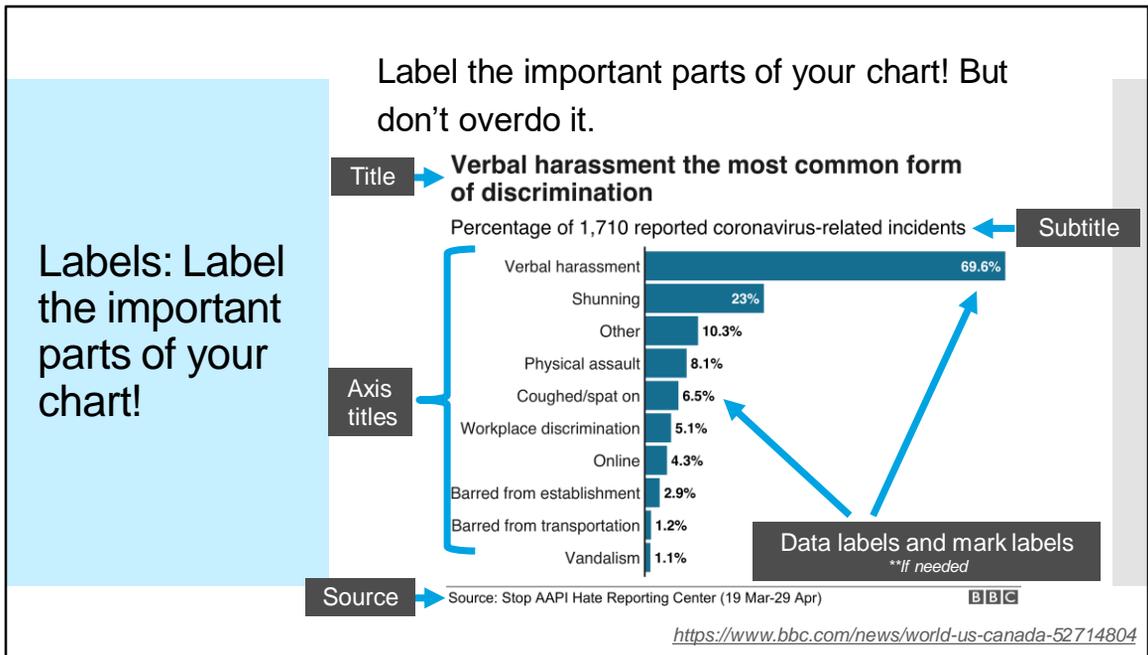
1. Label the important parts of your chart! Including a title, axes – and data marks as needed.
2. Use language that is appropriate for your audience.
3. Make sure it's understandable; consider how your visual will be delivered.

Labels are another key component of our graph which help interpretation. All of your labeling should really be in service to #3 – so that your chart is easily understood by your viewers.

For #1, Most importantly, this means having a title and axes on your chart. You may also want to add further labeling – such as data labels, or a subtitle. Again, be careful of overwhelming your audience with information – we want to keep the information that's critical at the forefront.

Another key element to consider is #2 – what language are you using? This includes thinking about the content of your language – is it too technical for your audience? Does it use jargon or abbreviations that people don't know? But also thinking about the language as well – is your audience going to be more comfortable viewing and interpreting the materials in Spanish? Another language?

Again all of this is in service to #3 – making sure the chart is understood



A recap of labels so you can see where each one fits:

Out title – what’s the main takeaway, or main content in the graphic

Our subtitle, which gives further detail *if needed*

Axis titles – this tells us what is happening in each row or column

Data labels – gives the specific value of each part of the visual

Source – where did the data come from

Remember – you don’t want to over clutter your chart and overwhelm the audience
You may not need data labels, or a subtitle – be sparing!

Chart from: <https://www.bbc.com/news/world-us-canada-52714804>

Labels: make the visual fit for the audience

Consider how your audience is consuming the visual, and their comfort level. Label your visual accordingly so your audience can interpret it!

- Are they looking at the visual on their own, or is someone explaining it to them?
- What is their level of comfort with data visualization?
- How often have they seen this visualization?
- How long will they have to look at this visualization?

As we said, our end goal is understanding.

The last thing I'd encourage you to think about with labeling is again the audience – how will they consume the visual?

Are they comfortable with visualization and don't need much guidance? Will you be available to explain to them the material?

Do they have a half hour to look at the visual and digest, or will they only have two minutes between meetings to glance at it?

Depending on the audience you may need more or less labeling to help guide them through the chart

Font: Some guidelines

1. Most important information is largest. Use a logical hierarchy when picking font size.
2. Make your font accessible. Consider how your visual will be delivered.

The last formatting piece to consider is font

Two key pieces here:

- 1) Make the font size logical – things that are the most important come first
- 2) Make it accessible. We talked about color blindness, and similarly we want to consider other possible visual impairments – how easily can someone read your text?

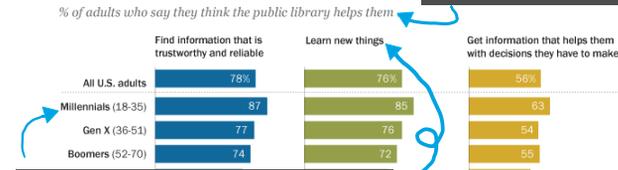
The sizing of your font should have a hierarchy that conveys relative importance!

Font: Most important text is the largest

The **title** should be largest

Millennials more likely than older generations to say trustworthy information, learn new things and make

Followed by the **subtitle**



Then the **headers** and **axis titles**

Source: Survey conducted Sept. 29-Nov. 6, 2016.
PEW RESEARCH CENTER

And finally the **data labels** and **mark labels**

https://www.pewresearch.org/fact-tank/2017/08/30/most-americans-especially-millennials-say-libraries-can-help-them-find-reliable-trustworthy-information/ft_17-08-22_libraries_generation/

The largest font should be the most important

So your title should be largest

The next most important should be the next largest

Typically this goes:

Subtitle

Headers / axis titles

Data labels / marks

Chart from: https://www.pewresearch.org/fact-tank/2017/08/30/most-americans-especially-millennials-say-libraries-can-help-them-find-reliable-trustworthy-information/ft_17-08-22_libraries_generation/

Font: Choose font that is accessible

Make sure to select font that is accessible for all audiences! Keep in mind variable levels of eyesight and the color of the font.

It's harder to read this...

...but it's not hard to read this!

A good rule of thumb is: minimum 9 point font for printed materials; minimum 20 point font for presented materials.

Accessibility is also key here

A small font, that is low contrast (like grey on white) is hard to read

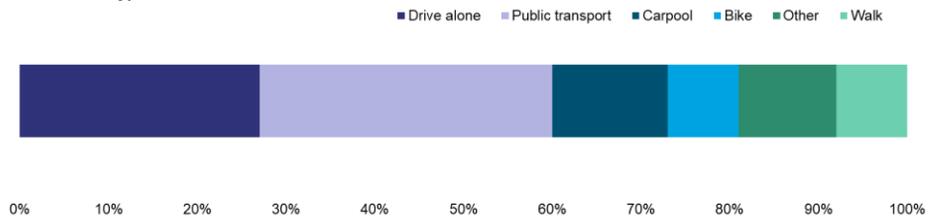
A big font that is high contrast (like black on white) is easy to read

A good rule of thumb is that printed materials are minimum 9pt
Presented are minimum 20 pt

You want to visualize how adults in Washington commute to the office.

In the following visual, what would you change in the formatting to make it more readable?

Figure 1. Commute type

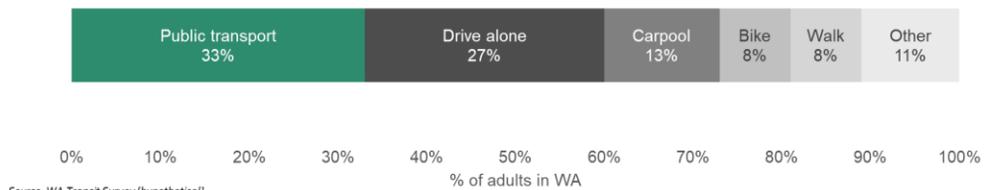


You want to visualize how adults in Washington commute to the office.

We changed the color scheme, re-ordered our data, integrated the legend, and added labels and a title.

How do adults in Washington commute to work?

Most people take **public transit**, but a substantial number still drive alone.



Source: WA Transit Survey [hypothetical]

A checklist for formatting your visual

A handy [printable checklist](#) summarizes the formatting you should consider:

- Descriptive title in the upper left corner
- Subtitle and/or annotations to provide further information
- Text size is hierarchical and readable
- Text is horizontal
- Data are labeled correctly
- Labels are used sparingly
- Proportions are accurate
- Data are accurate
- Data are intentionally ordered or sorted
- Axis intervals are equidistant
- Graph is two-dimensional
- Display is free from decoration
- Color scheme is intentional
- Color is used to highlight key patterns
- Color is legible when printed in black and white
- Color is legible for people with colorblindness
- Text sufficiently contrasts with background
- Gridlines, if present, are muted
- Graph does not have border line
- Axes do not have unnecessary tick marks or axis lines
- Graph has one horizontal and one vertical line
- Graphs highlight significant finding or conclusion
- The type of graph is appropriate for the data
- Graph has appropriate levels of precision
- Individual chart elements work together to reinforce the takeaway message

https://datavizchecklist.stephanieevergreen.com/assets/DataVizChecklist_Feb2018.pdf

A cheat sheet for formatting is available here, and is a good way to review / remind yourself of all the formatting pieces needed

https://datavizchecklist.stephanieevergreen.com/assets/DataVizChecklist_Feb2018.pdf

Agenda

- 1 Why visualize your data?
- 2 A framework for visualization
- 3 Purpose
- 4 Content
- 5 Structure
- 6 Formatting
- 7 **Bringing it all together**

A case study
of the life of a
visual from
start to finish

Purpose:

- Why am I creating this visual?

Content:

- What data am I using, and what matters?

Structure:

- How can we best present our data?

Formatting:

- What else is needed to make our graphic visually accessible?

Audience:

- *Who is our user and what do they need?*

Case study: Audience

A hypothetical case study:

We just finished a project that offered parenting and childcare support to families.

We need to present the results of our program to:

- Donors
- County officials

Case study: Purpose

There are two program activities offered monthly:

- Play groups for young children
- Parenting skill classes

Purpose: How often did families participate in each of the program activities in the last 12 months?

Case study: Content

We reviewed the program logs for the 188 of the program participants to find out how often they participated in each activity.

We discovered four attendance 'categories' :

- Never
- Once
- Twice
- Three or more times

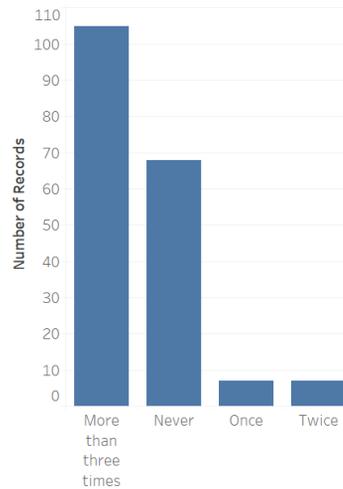
We have our purpose: How often did families participate in each of the program activities in the last 12 months?

We have our content: program logs for 188 participants for their attendance, broken into four attendance categories

So now we want to think about structure

We're fundamentally comparing different groups (the different attendance categories) so we want to use a bar chart

Case study: Structure



Purpose: How often did families participate in each of the program activities in the last 12 months?

This is okay, but there are some limits here:

No title, so we're not sure which program activity

The sorting of the columns is high to low, but it jumps from 3 times to never

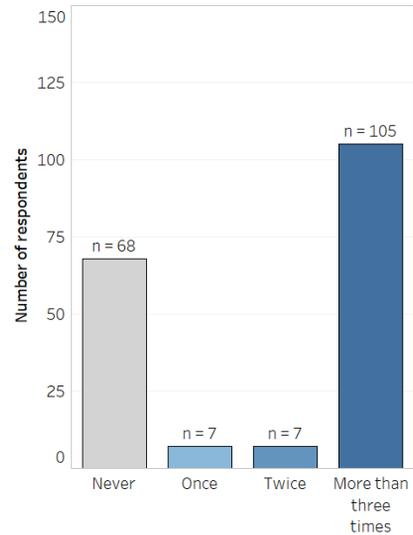
It's not particularly visually appealing

Case study: Formatting

To improve legibility:

- Ordered the columns from least to greatest attendance
- Added title and subtitle
- Added axis labels
- Added data labels which show the 'n' of each answers
- Added colors that indicate the frequency

Most program participants attended a play group multiple times
Frequency of participation



Clean it up by ordering the columns by attendance numbers

Added title and axis labels

Add colors which help indicate frequency (light to dark)

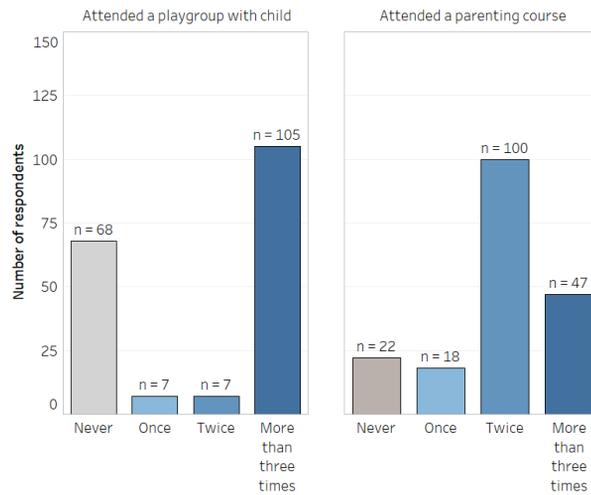
Now, we had two program activities: one for parents and one for children

This only shows one program activity. Can we show both?

Bonus: Small multiples?

Most program participants attended a class and play group multiple times

Frequency of participation in program activities



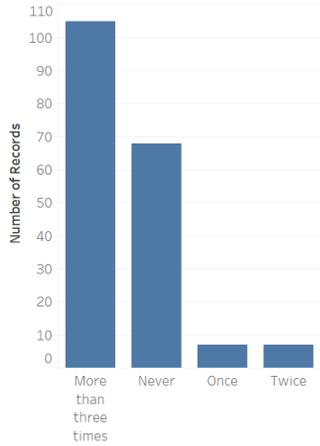
Source: data is generated for purposes of example / anonymity

We asked participants the same type of question for the two program activities and have the same attendance categories for both

So we used small multiples to answer this question!

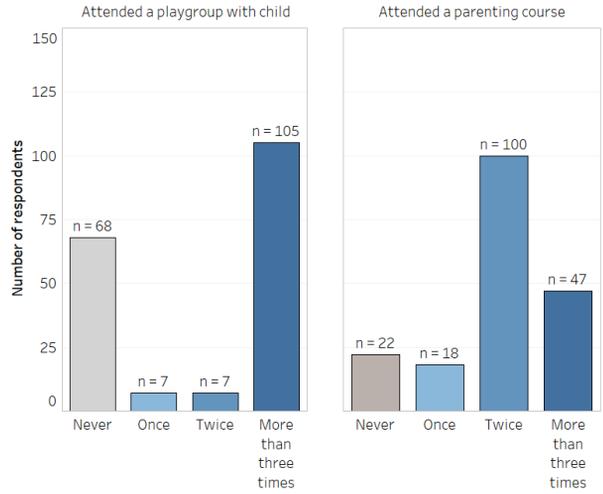
The final visual!

How often did you attend a parenting skills course?



Most program participants attended a class and play group multiple times

Frequency of participation in program activities



Source: data is generated for purposes of example/ anonymity

But what tools
do I need to
create these
visuals
myself?

There are lots of different software you can use to visualize! They do not need to be expensive, or hard to use.

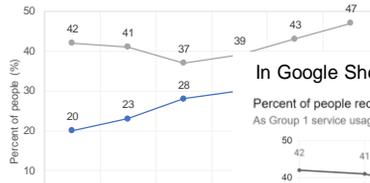
Some possibilities:

- Excel
- Google Sheets
- Tableau
- Many more...

All of these
can create
beautiful
visuals!

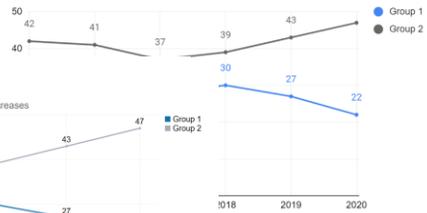
In Excel:

Percent of people receiving a service from 2015-2020
As Group 1 service usage increase, Group 2 service usage decreases



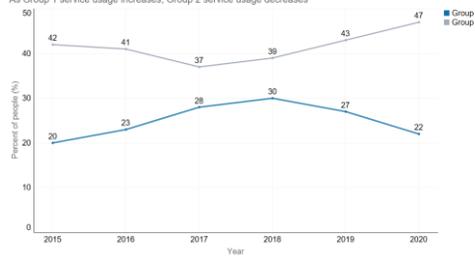
In Google Sheets:

Percent of people receiving a service from 2015-2020
As Group 1 service usage increases, Group 2 service usage decreases



In Tableau:

Percent of people receiving a service from 2015-2020
As Group 1 service usage increases, Group 2 service usage decreases



Excel Resources

Excel can make all sorts of graphs! Some resources which include how-to's and templates:

- [Juice Analytics](#)
 - Downloadable Excel templates
- [Depict Data Studio](#)
 - How-to's in Excel for Data Viz
 - Note: you have to click through a series of links, but they are there!
- [Stephanie Evergreen](#)
 - How-to's in Excel for Data Viz!

<http://labs.juiceanalytics.com/chartchooser/index.html>

<https://depictdatastudio.com/charts/>

<https://stephanieevergreen.com/how-to/>

Google Resources

Google Sheets can also make all sorts of graphs! Some resources which include how-to's:

- [Google Support](#)
 - Shows you all the different types of charts available in Google Sheets and gives a link to a how-to for each chart type. Some common ones:
 - [Line graph](#)
 - [Bar chart](#)
 - [Pie chart](#)

<https://support.google.com/docs/answer/190718?hl=en>

<https://support.google.com/docs/answer/9142593>

<https://support.google.com/docs/answer/9142829>

<https://support.google.com/docs/answer/9143036>

Tableau Resources

If you are interested in using Tableau:

- [Tableau Public](#) allows you a free license, *but then your visuals are all publicly available.*
 - Not good for sensitive data, or data that has personal information attached!
- [Tableau for Good](#) gives free licenses to non-profits with annual operating budgets under 5 million;
 - To get your application process and download, there is an administrative fee of ~\$50
- Tableau has training resources online, including [free videos](#) and [90 days of free e-learning courses](#).
 - Note: e-learnings are not normally free!

<https://public.tableau.com/en-us/s/>

<https://www.tableau.com/foundation/license-donations>

<https://www.tableau.com/learn/training/20202>

https://www.tableau.com/learn/training/elearning?utm_campaign=2017049_EGCore_CORPR_USCA_en-US_2020-04-16-Cust-COVID-Company-Statement&utm_medium=Email&utm_source=Eloqua&domain=path.org&eid=CTBLS000010157945&elqTrackId=180c196ab562479d807fbaf410f47219&elq=8af59885f07344f9b0693f2fa5f4e53b&elqaid=42709&elqat=1&elqCampaignId=41953

Many more?

Chartmaker [aggregates 'how-tos'](http://chartmaker.visualisingdata.com/) for all different types of platforms!

	Amazon QuickSight	ArcGIS	ChartJS	Charticulator	D3.js	Data Illustrator	Datavrapper	Flourish
Bar chart	●				●●●●	○	●●●●	○
Clustered bar chart	●				●	○	●●●●○	○
Bullet chart				●	●		●●	
Waterfall chart				●	●			
Radar chart			○		●			

<http://chartmaker.visualisingdata.com/>

<http://chartmaker.visualisingdata.com/>

Question: what about infographics?

“Infographic” describes a visual communication which includes data, text, or both.

- You can use all sorts of graphics in an infographic!
- Tell a story with your infographic; all the pieces should be part of a narrative
- Make it easy to navigate the story using thoughtful design
- Don't over rely on text; don't rely on data. Balance both to tell the story!
- Continue to adhere to good data visualization and graphic design practices

Some resources:

- [What is an infographic?](#)
- [Bad infographics and some principles](#)
- [Storytelling with infographics](#)
- [Flourish- Free data visualization platform](#)

<https://www.columnfivemedia.com/infographic>

<https://stephanieevergreen.com/bad-infographics/>

<http://www.storytellingwithdata.com/blog/2020/14/storytelling-with-infographics>

<https://flourish.studio/>

How can you request assistance, or ask questions about data and evaluation?



**Communities
Count**

Free Data Consultation & Technical Assistance
Email: communitiescount@kingcounty.gov

**See more data and trainings, visit us at:
www.communitiescount.org**

Help us learn
and improve!

Before you leave this topic, please help
us better meet your needs by
completing this short feedback form:

[https://www.surveymonkey.com/r/cc
dataworkshop](https://www.surveymonkey.com/r/ccdataworkshop)

What questions and reactions do you have?