Intro to Visualization

Lesson Objectives

After this lesson, you will be able to…

- Describe why data visualization is important for communicating results.
- Identify how to select the correct visualization to use based on the data being presented.
- Identify characteristics to clearly communicate through data visualizations.

How Do we Make Sense of a Data Set?

We're only looking at 1/3 of this data set! While all the data we need is here, it is difficult to make sense of and draw any meaning from.

Airport	Reason for Delay	Number of Operations	% of Total Operations	Delayed Minutes	% of Total Delayed Minutes
SFO	**Air Carrier Delay**	28,899	4.85%	1,970,571	30.57%
	Aircraft Arriving Late	37,618	6.31%	2,565,218	39.79%
	Security Delay	165	0.03%	6,716	0.10%
	National Aviation System Delay	34,642	5.81%	1,628,412	25.26%
	Extreme Weather	2,482	0.42%	276,076	4.28%
	Total Operations	596,046	100.00%	6,446,993	100.00%
ATL	**Weather**	387	27.98%	13,939	26.72%
	Volume	948	68.55%	36,505	69.99%
	Equipment	0	0.00%	0	0.00%
	Closed Runway	47	3.40%	1,665	3.19%
	Other	1	0.07%	52	0.10%
	Total Operations	1,383	100.00%	52,161	100.00%
DEN	**Weather**	545	66.54%	25,402	76.76%
	Volume	221	26.98%	6,199	18.73%
	Equipment	0	0.00%	0	0.00%
	Closed Runway	49	5.98%	1,376	4.16%
	Other	4	0.49%	115	0.35%
	Total Operations	819	100.00%	33,092	100.00%

So What Is Data Visualization?

- A quick, easy way to convey concepts that from from large data sets.
- We can use these charts, graphs, or illustrations to visualize large amounts of complex data.

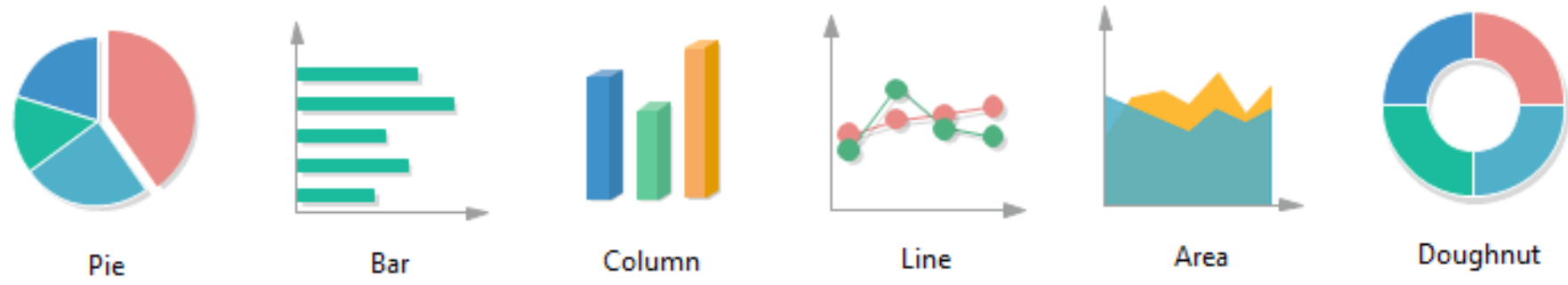
Criteria for Crafting a Good Visualization

- Visualizations should follow three (plus one) rules. They should be:

1. Simplified
2. Easy to Interpret
3. Clearly Labeled
4. (Bonus) Interactive

How Do you Choose the Right Chart Type?

With so many chart types, it can be difficult to know how best to display your data.



When creating a visualization first think about the variables you are showing (words, categories, numbers, etc., the volume of data, and the central point you are hoping to communicate through your visualization.

The Bar Chart in Action

Looking at this bar chart, what do you notice about this visualization?

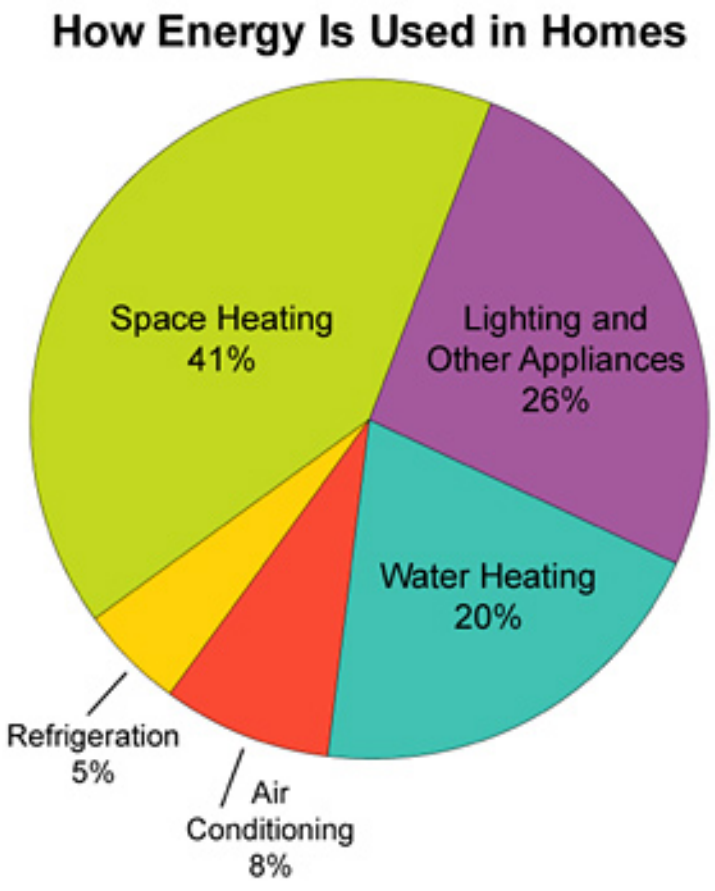
When to Use a Bar Chart

Bar charts are one of the most simple and frequently used chart types. They are useful for illustrating either one string or one numeric variable, quickly comparing information, or for show exact values.

When thinking about using a bar chart consider: - Will you use vertical or horizontal bars? - How will you number your axis (it is always best to start at zero)? - How will you order your bars?

The Pie Chart in Action

As you can see from this example pie charts can be effective for proportions or percentages.



Source: U.S. Energy Information Administration

When to Use the Pie Chart Type

Pie charts are commonly misused. They show a part-to-whole relationship when the total amount is one of your variable and you'd like to show the subdivision of variables.

When thinking about using a pie chart consider: - The more variables you have, as in the more slices of your pie you'll have, the harder it is to read. - Area is *very* difficult for the eye to read, so if any of your wedges are similarly sized think about a different chart type. - If you want to compare data, leave it to bars or stacked bars. If your viewer has to work to translate pie wedges into relevant data or compare pie charts to one another, the key points you're trying to convey might go unnoticed.

The Scatter Plot in Action

This scatter plot uses a combination of text, coloring, and labelling to describe the data. What is clear or unclear from this chart about the data set?

When to Use a Scatter Plot

Scatterplots are great for data dense visualizations and clusters. They are most effective for trends, concentrations, and outliers. They can be especially useful to see what you want to investigate further.

When thinking about using a scatter plot consider: - This chart type is not as common so can me more difficult for an audience to read. - If dots are covering up each other, consider a different chart type. - A bubble chart is one variation on the scatter plot.

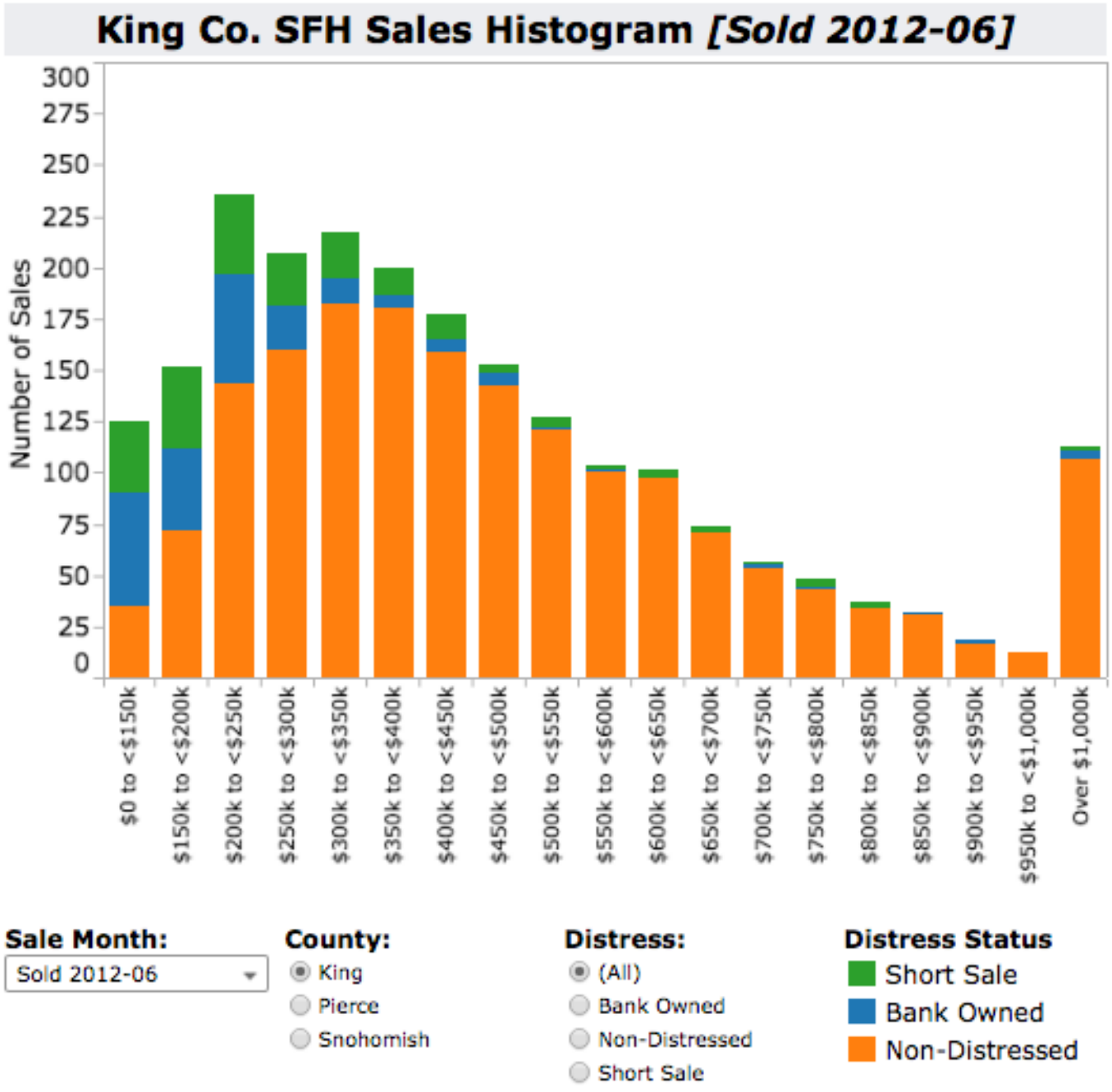
Knowledge Check: Choosing a Chart

Annual sales in each state for a grocery store chain?

- Bar chart.
- Pie chart.
- Scatterplot.

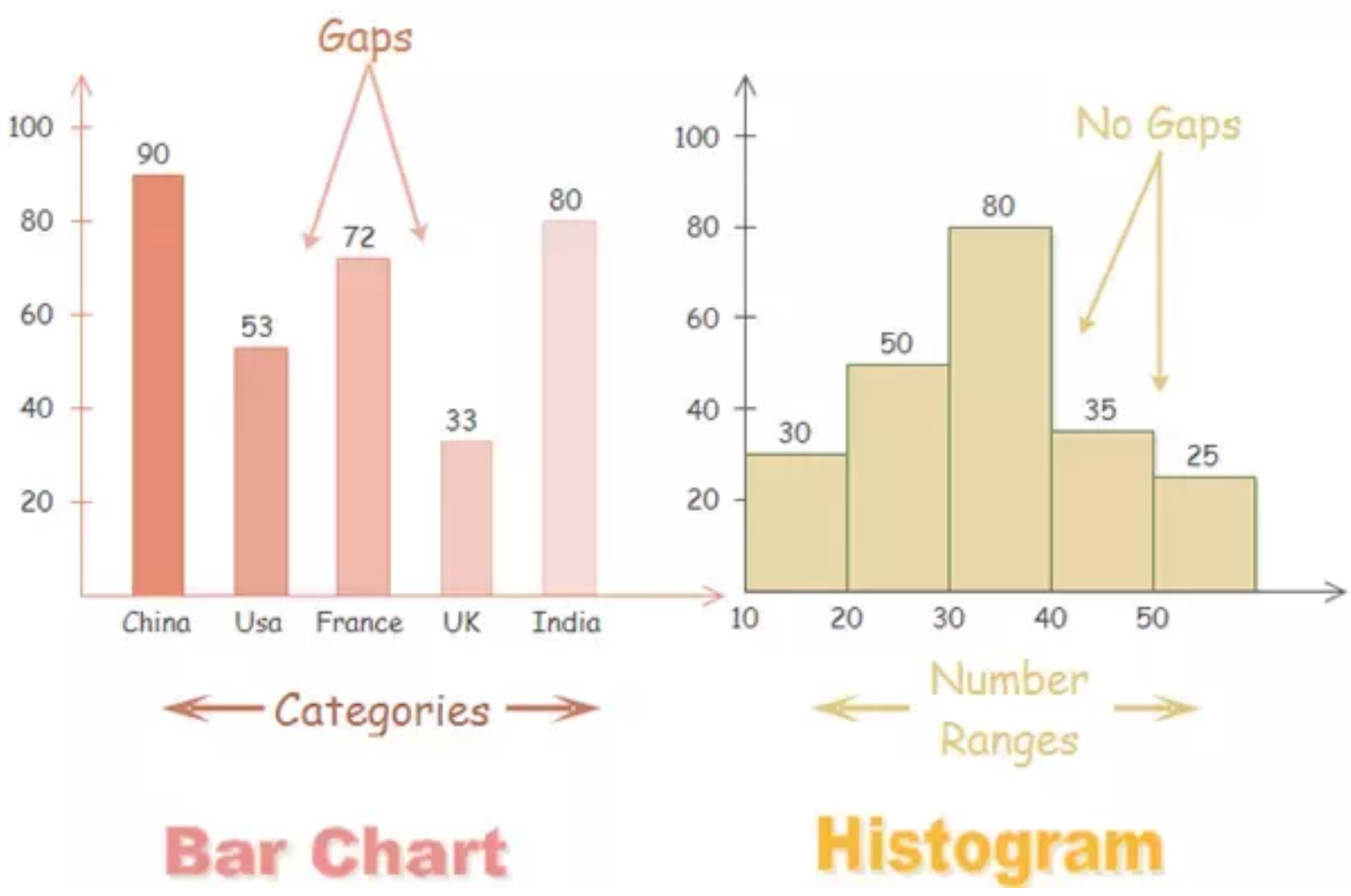
When to Use a Histogram

- Effective for distribution across groups.



Bar Chart vs Histogram

The main difference between a bar chart and histogram is that histograms are used to show distributions of variables while bar charts are used to *compare* variables.



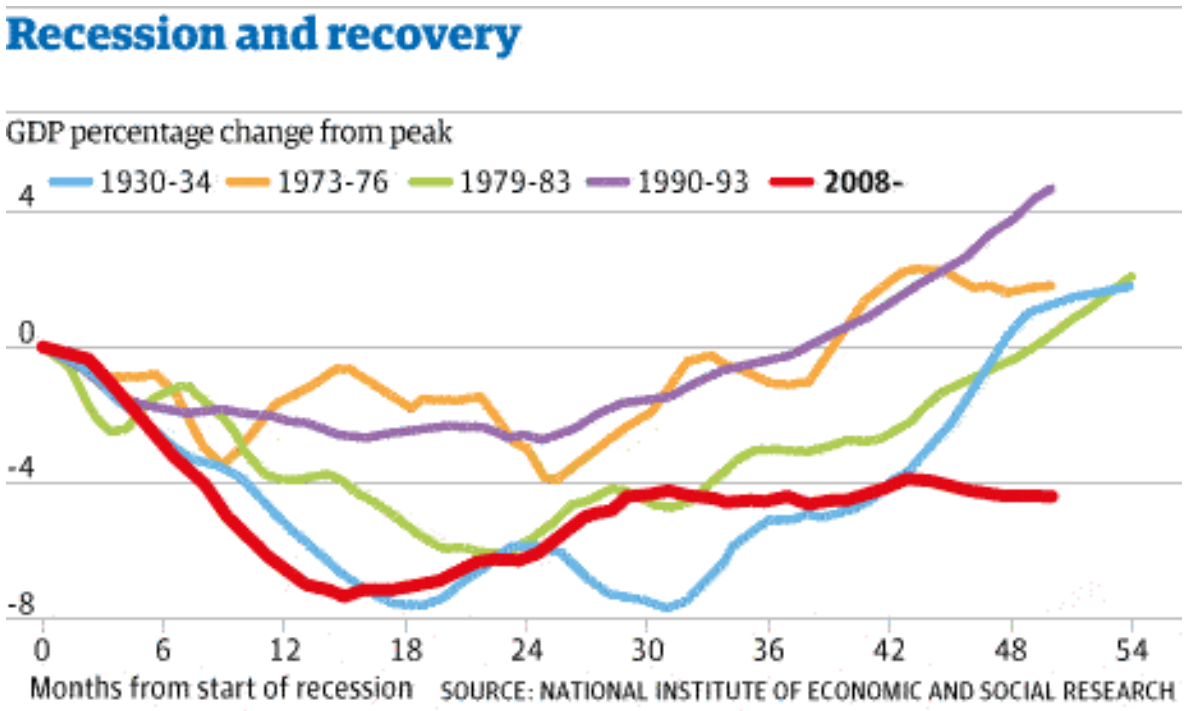
Which type of chart?

Relationship of average income to education level?

- Bar chart.
- Pie chart.
- Scatterplot.
- Histogram.

A Line Chart in Action

Line graphs are an excellent way to show change over time. While bar charts can also show time, they don't show it in a continuous way like a line chart.



When to Use a Line Chart

Line charts are particularly good at showing how a variable change over time. They work best if you have one date variable and one number variable.

When thinking about using a line chart consider: - How many lines you'll need on your graph, the more overlapping lines there are, the harder your chart will be to read. - Consider how many colors you need to use for your lines. Giving each line its own color forces the viewer to scan back and forth from the key to the graph. - Individual data points can be hard to read, but line charts are good for showing overall trends. - Similar to bar charts, try and start at 0 on your x axis.

Knowledge Check: Which type of chart?

Change in average income since 1960 for American adults?

- Bar chart.
- Pie chart.
- Scatterplot.
- Histogram.

Returning to How to Choose the Right Chart

Check out this series of charts: `https://i.redd.it/e7alp8yrnb711.png`

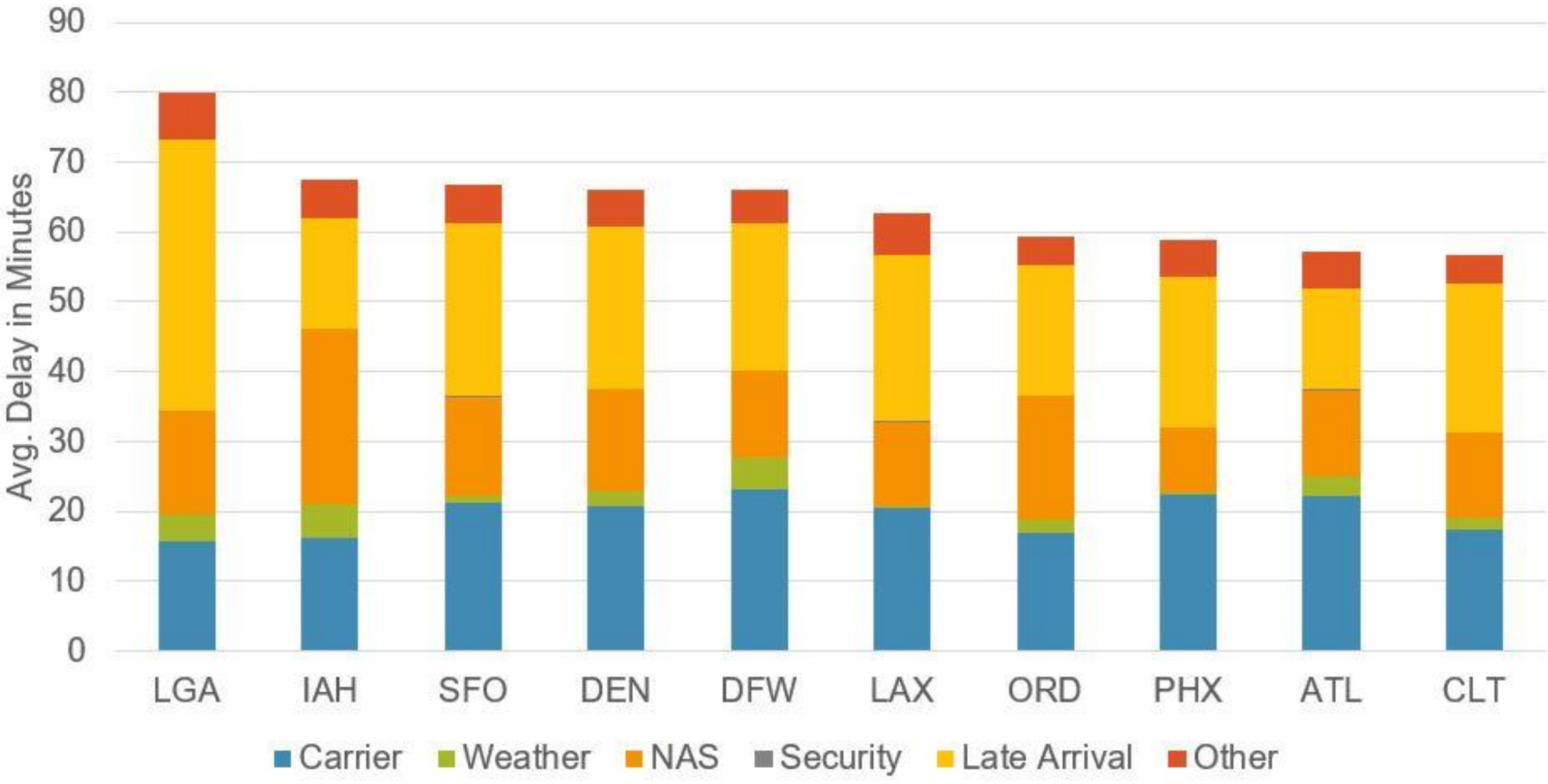
- Which is easiest to view the data?

None right! Choosing a chart type depends firstly, on the data you have. Secondly, on the clearest way to convey your message, and how these two aspects align.

Charts & Code

There is an increasing array of libraries and tools to allow us to use code to create visualize data in compelling and approachable ways.

Check out this complex chart that was made using Python!



Source: u/dx034 on Reddit

Group Activity: Exploring Good Visualizations

Get in small groups of 2-3.

[Go to https://www.reddit.com/r/dataisbeautiful/top/. These are all data visualizations created by people like you!

Pick one that you think is particularly good and one that is particular bad. Why? What are the characteristics?

Visual Attributes of Good Data Visualization

Some attributes affect our brain more strongly.

In order of focus: - Position - Color - Size

Summary

- The chart type you select should accurately represent the variables you are pulling from data in a way that is clearly readable for your audience.
- Visual considerations include: position, color, order, size. What else?
- With data visualizations becoming increasingly popular, a clean and clear chart goes a long way in conveying a message from a data set.

Additional Resources

- A great short article on when pie charts are better?
- A gallery of charts
- A Stats Video
- SAS: Data Viz
- A guide to when to use each chart
- 44 Types of Graphs
- Advice on making good visualizations
- Reddit's Data Visualization forum