Data Visual Design and Storytelling

Participant Guide

Table of Contents

Class activities 3

- Activity: data types 4
- Activity: anatomy of a chart 6
- Activity: class survey – part I 8
- Activity: common charts and graphs 9
- Activity: class survey – part II 13
- Activity: class survey – part III 18
- Activity: from insight to outcome 24
- Activity: reverse storyboarding 29

Additional resources 31

- Tool selection checklist 32
- Data visualization checklist 33
- Where to learn more 35

Class activities

Activity: data types

The Executive Summary from the World Bank's April 2020 Commodity Markets Outlook is presented below.

Read the summary and answer the questions on the following page.

Executive Summary

Almost all commodity prices saw sharp declines during the past three months as the COVID-19 pandemic worsened. Mitigation measures have significantly reduced transport, causing and unprecedented decline in demand for oil, while weaker economic growth will further reduce overall commodity demand. Crude oil prices are expected to average $35/bbl this year and $42/bbl in 2021—a sharp downward revision from October in both years. Non-energy prices are also expected to fall this year. Metals are projected to decline more than 13 percent in 2020, before recovering in 2021 while food prices are expected to be broadly stable. The risks to the price forecasts are large in both directions and depend on the speed at which the pandemic is contained and mitigation measures are lifted. [This document] investigates the impact of COVID-19 on commodity markets and compares it with previous disruption episodes. It finds that the impact of COVID-19 has already been larger than most previous events and may lead to long-term shifts in global commodity demand and supply. [This document] examines the impact of international commodity production agreements, with a particular focus on OPEC, and concludes that OPEC+, the last remaining international agreement to manage supply, is subject to the same forces that led to the collapse of its predecessors.

1. What examples of quantitative data are included in the summary that you read?

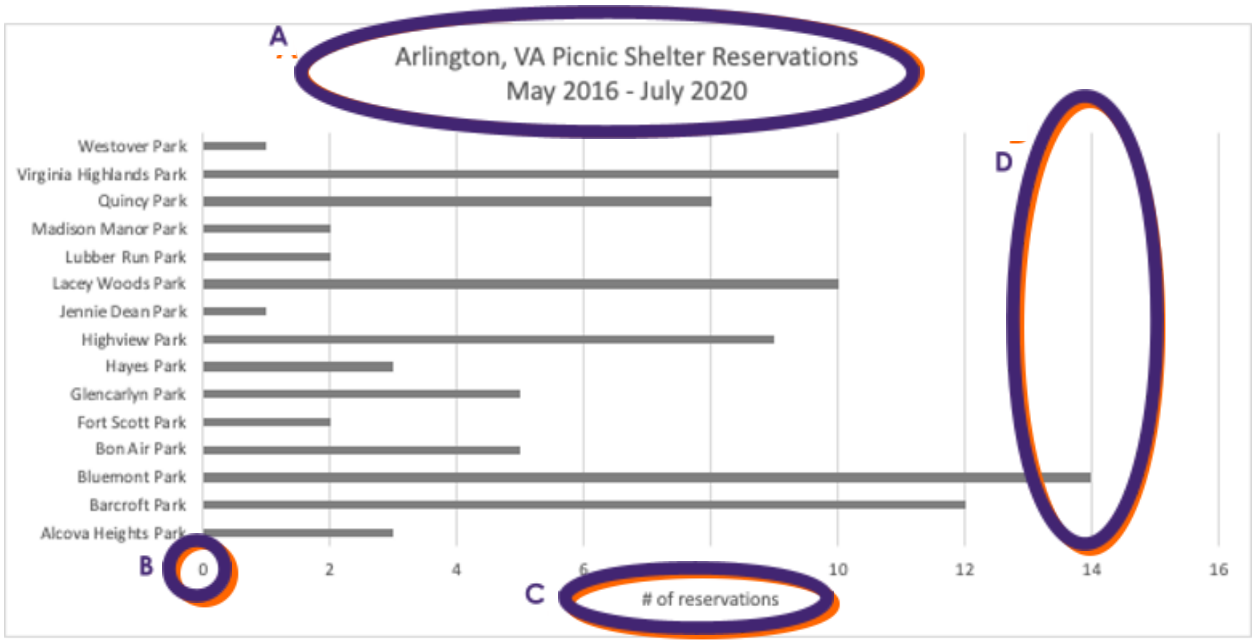
2. Do you see any examples of qualitative data in the summary? If so, what are they?

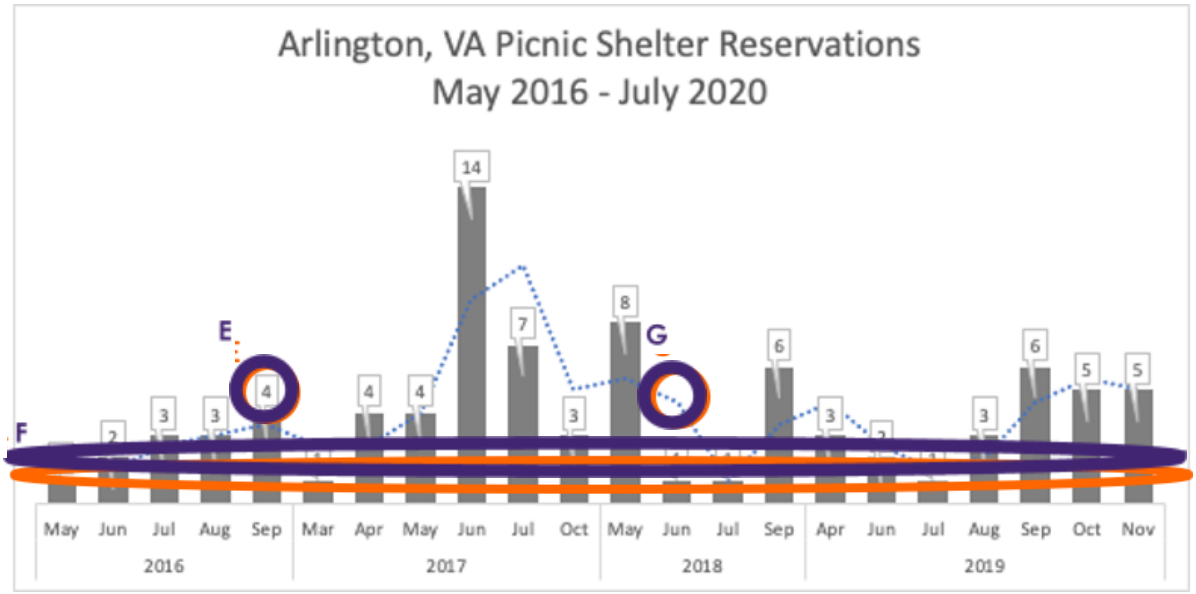
3. Do you see examples of continuous data are mentioned in this summary? If so, what are they?

4. What types of visualizations might make this summary easier to digest?

Activity: anatomy of a chart

Review the following charts and then name the elements in the table on the next page.





Choose the correct element from the pull-down menu.

Element	Name
A	trendline
B	trendline
C	trendline
D	trendline
E	trendline
F	trendline
G	trendline

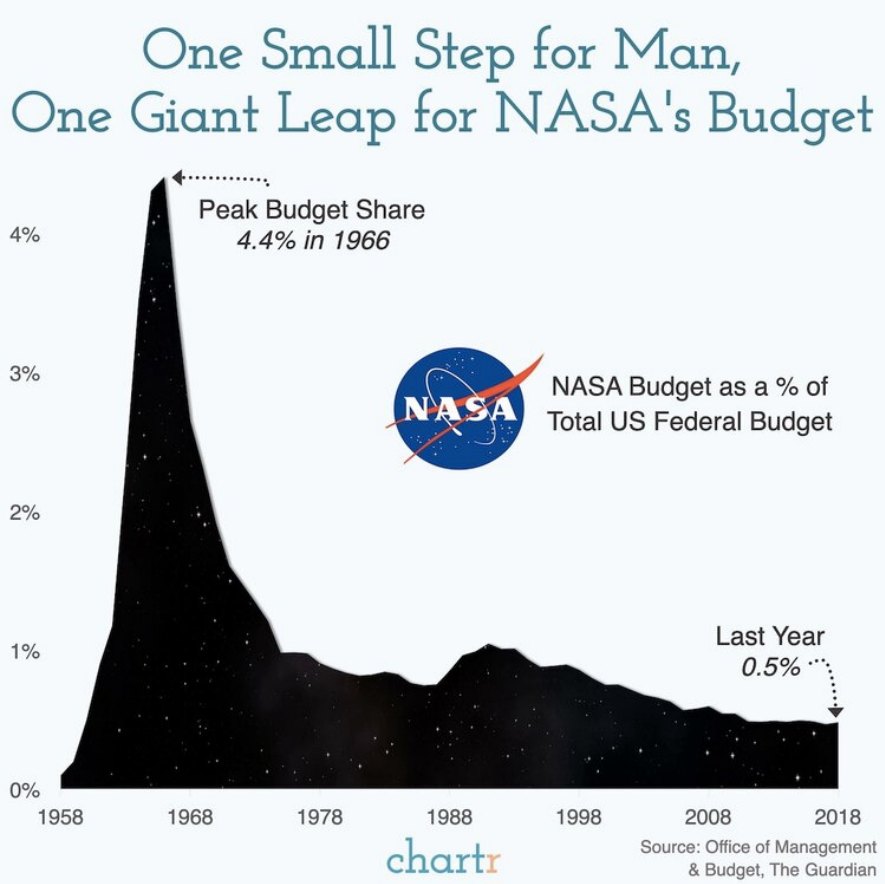
Activity: class survey – part I

Please complete the class survey found at:

https://www.surveymonkey.com/r/MCJNZQ6

Activity: common charts and graphs

Chart 1
Review the chart below and answer the two questions that follow.



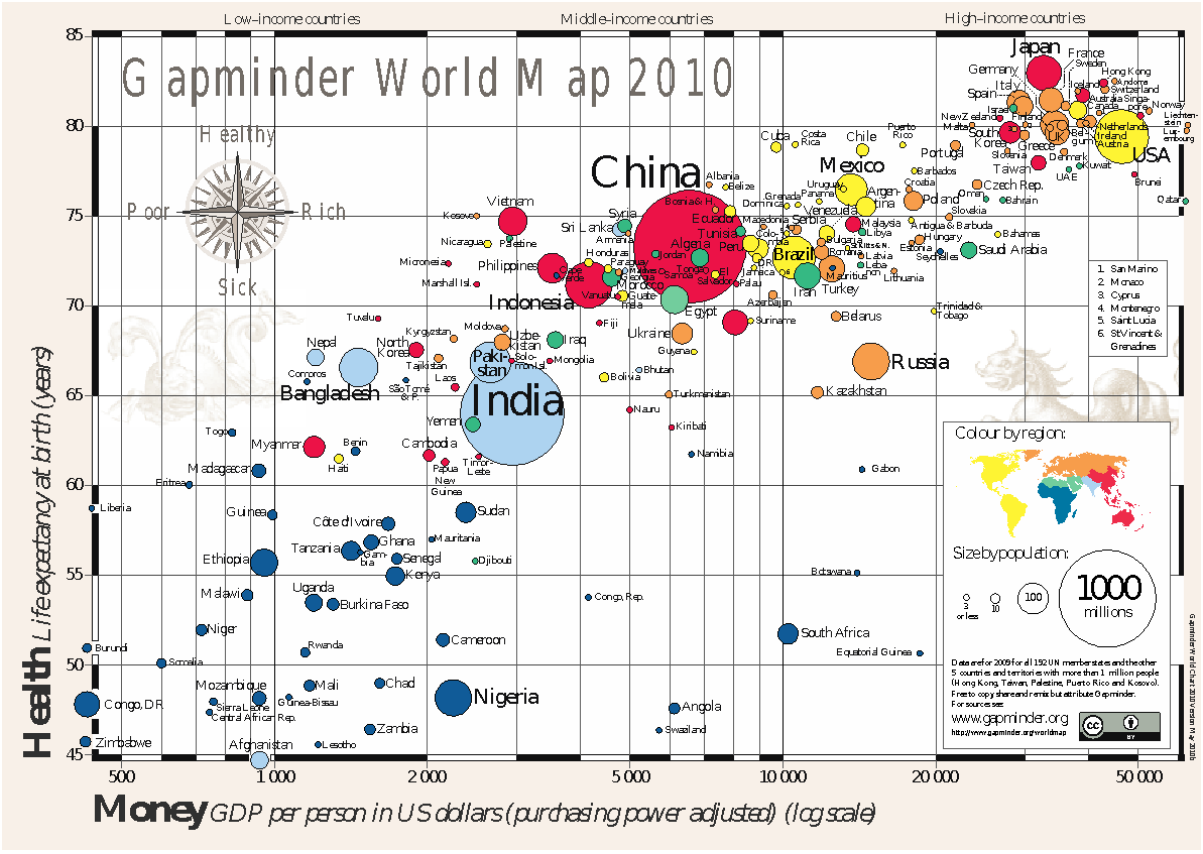
1. What type of chart is this?

a. Line chart
b. Area chart
c. Scatter plot
d. Histogram

1. What type of data is presented in this chart?

a. Discrete, qualitative
b. Discrete, quantitative
c. Continuous, qualitative
d. Continuous, quantitative

Chart 2
Review the chart below and answer the two questions that follow.



Source:
https://static.gapminder.org/GapminderMedia/wp-uploads/pdf_charts/GWM2010.pdf

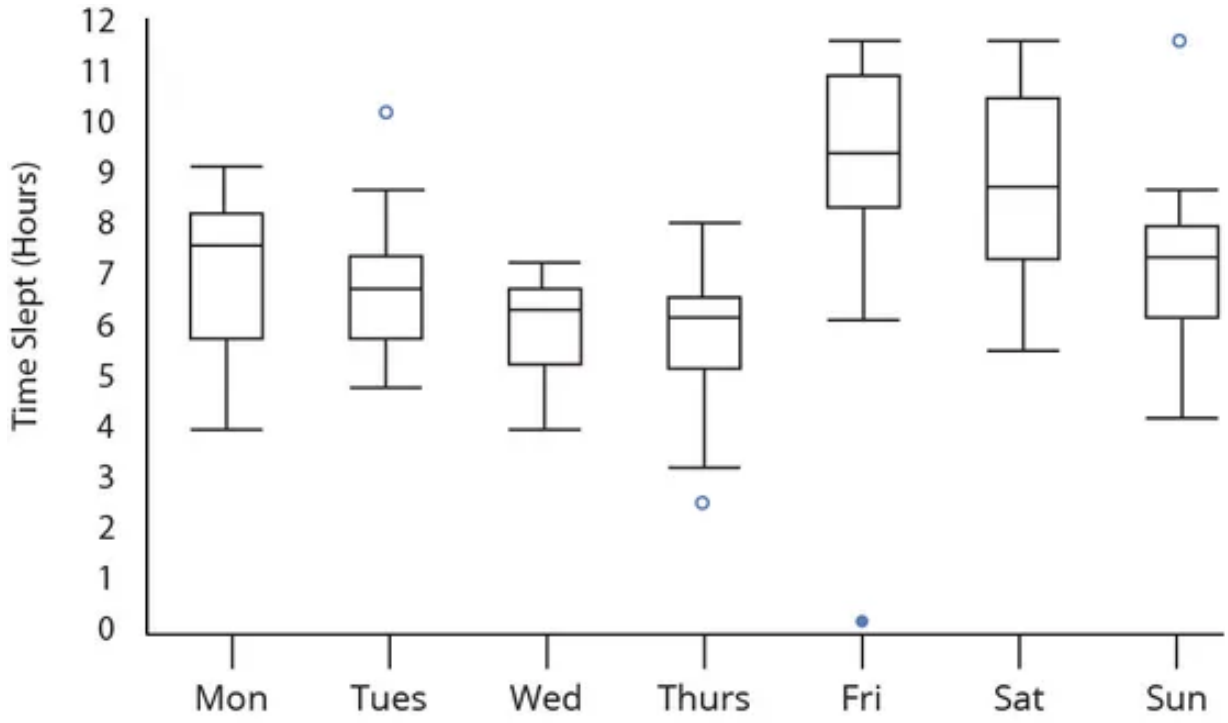
1. What type of chart is this?

a. Scatter plot
b. Bubble chart
c. Heatmap
d. Boxplot

1. What does the size of the bubbles represent?

a. Expected years of life at birth
b. GDP per person
c. The population of each country
d. The size has no significance

Chart 3
Review the chart below and answer the three questions that follow.



Source: https://www.simplypsychology.org/boxplots.html

1. What type of chart is this?

a. Scatter plot
b. Bubble chart
c. Heatmap
d. Boxplot

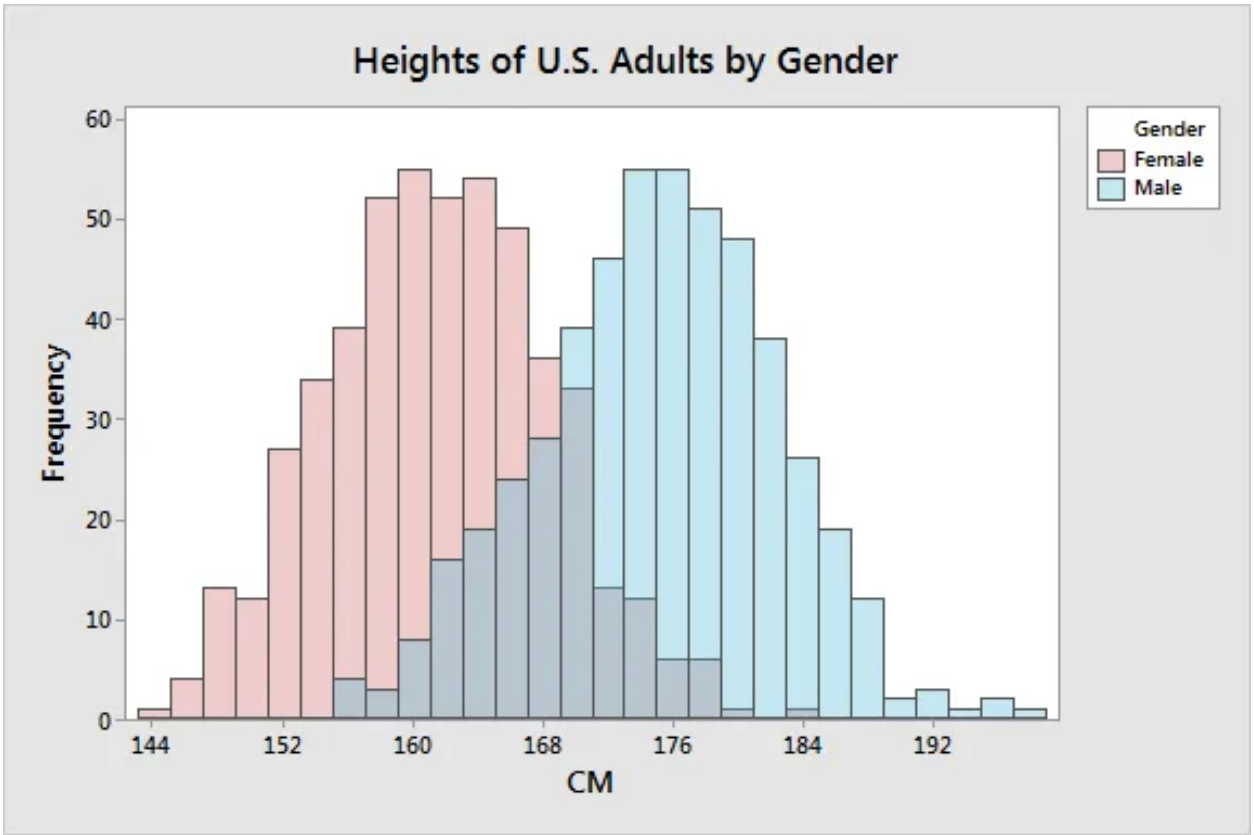
1. Which of the following days sees the most variance in number of hours slept?

a. Tuesday
b. Wednesday
c. Sunday
d. Friday

1. The data for which of the following days includes an outlier?

a. Monday
b. Wednesday
c. Saturday
d. Friday

Chart 4
Review the chart below and answer the three questions that follow.



Source: https://statisticsbyjim.com/basics/histograms/

1. What type of chart is this?

a. Bar chart
b. Line chart
c. Histogram
d. Boxplot

1. **What type of data is represented by the bins?**

a. Discrete, qualitative
b. Discrete, quantitative
c. Continuous, qualitative
d. Continuous, quantitative

1. What type of data do the two colors represent?

a. Discrete, qualitative
b. Discrete, quantitative
c. Continuous, qualitative
d. Continuous, quantitative

Answers: 1.b, 2.d, 3.b, 4.c, 5.d, 6.d, 7.d, 8.c, 9.d, 10.a

Activity: class survey – part II

Review the questions from the class survey that you all took earlier, which are presented in the table below.

Question	Multiple choice options
How many years have you worked for your current employer?	*<1* *1-5* *6-10* *11-19* *20+*
How confident are you in your ability to describe the difference between a histogram and a boxplot?	*Extremely confident* *Very confident* *Somewhat confident* *Not so confident* *Not at all confident*
Roughly what percentage of your time at work is spent collecting, analyzing, or visualizing data?	*N/A – short answer*
When was the last time you created a chart or graph to present data?	*In the last week* *More than a week ago but less than a month ago* *More than a month ago*
Are you interested in learning the Python programming language?	*Extremely interested* *Very interested* *Somewhat interested* *Not so interested* *Not at all interested*
What time did you get out of bed today?	*N/A – short answer*
What types of pets do you own?	*Dog(s)* *Cat(s)* *Fish* *Birds(s)* *Other (please specify)*

How many dogs do you own?	*0* *1* *2* *3* *4 or more*
How old are you?	*Under 18* *18-24* *25-34* *35-44* *45-54* *55-64* *65+*
Would you ever go sky diving?	*Definitely would* *Probably would* *Probably would not* *Definitely would not*

Think about what story you could uncover or tell with the results and who you want to tell it to.

On the following page, describe, draw, or create a quick visualization to express how you'd like to see that story presented.

Use this space to describe, draw, or otherwise create a quick visualization that tells a story about the survey data.

Activity: class survey – part III

Review the results of the class survey that you all took earlier at:
https://www.surveymonkey.com/stories/SM-WY5Y5RS2/

Use the tables below to jot down your notes about how you'd improve the visuals provided.

Visual 1

How many years have you worked for your current employer?	
Chart type	
Visual design	
Clutter	
Other	

Visual 2

How confident are you in your ability to describe the difference between a histogram and a boxplot?	
Chart type	

Visual design	
Clutter	
Other	

Visual 3

Roughly what percentage of your time at work is spent collecting, analyzing, or visualizing data?	
Chart type	
Visual design	
Clutter	
Other	

Visual 4

When was the last time you created a chart or graph to present data?	
Chart type	

Visual design	
Clutter	
Other	

Visual 5

Are you interested in learning the Python programming language?	
Chart type	
Visual design	
Clutter	
Other	

Visual 6

What time did you get out of bed today?	
Chart type	
Visual design	

Clutter	
Other	

Visual 7

What types of pets do you own?	
Chart type	
Visual design	
Clutter	
Other	

Visual 8

How many dogs do you own?	
Chart type	
Visual design	
Clutter	

Other	

Visual 9

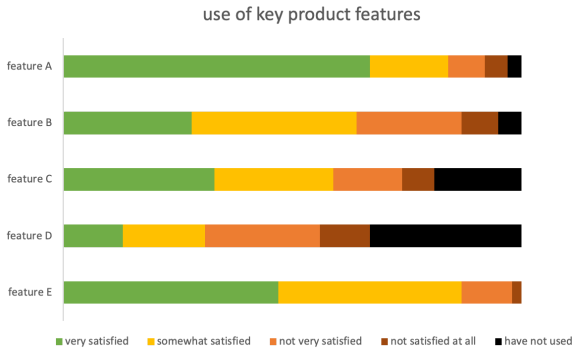
How old are you?	
Chart type	
Visual design	
Clutter	
Other	

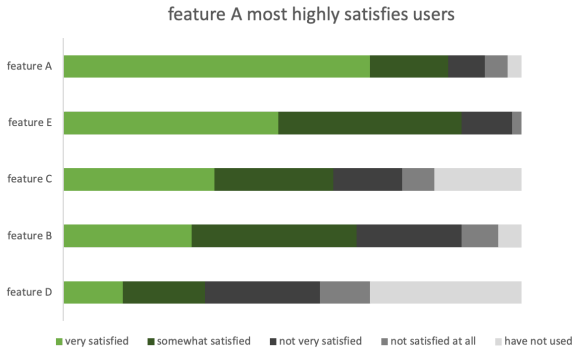
Visual 10

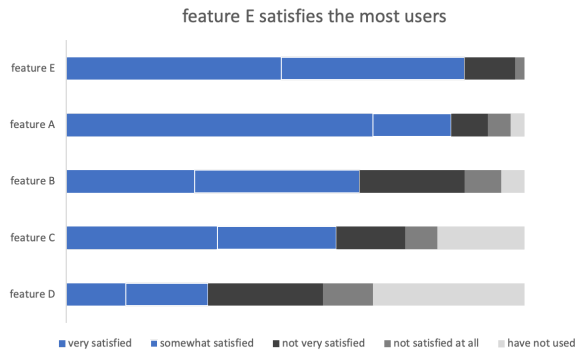
Would you ever go sky diving?	
Chart type	
Visual design	
Clutter	
Other	

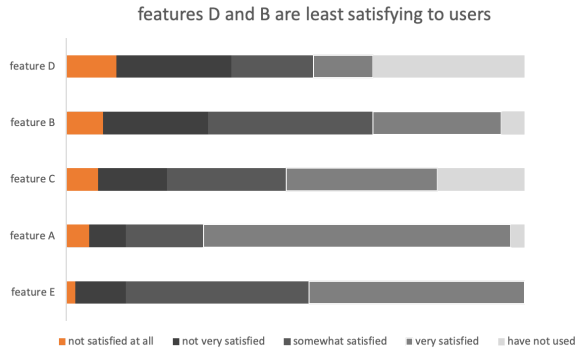
Activity: from insight to outcome

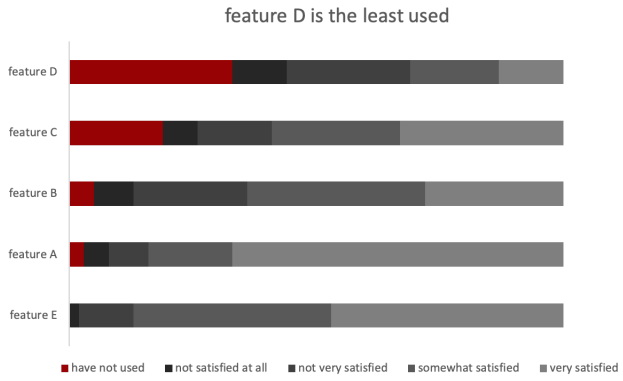
Recall the insights generated from the results of the user satisfaction survey about the **online banking app**, which we reviewed in class.











Now, suppose the purpose of that survey was to help determine which feature to retire in the next major app update. Each feature corresponds to a different contact method:

A. schedule an appointment
B. direct message a representative
C. tweet
D. email a representative
E. request a phone call

Based on the data, you have generated the following insight: **"Email a representative" should be retired, since it is the least used feature by a wide margin.** You have also determined that **many users prefer "direct message a representative," which performs a similar function as the "email a representative" option.**

You are going to present this information to two different audiences.

Your first audience is the **marketing team**. Their main concern is to generate excitement for the app by creating a bold and flashy campaign about the latest redesign. They want to begin development quickly, but they tend to worry about users who will be turned away by changes to the product. They are fairly knowledgeable about polling and surveys, and they expect you will focus on how to balance what users want with what is most new and innovative about the next generation of product.

Your second audience is the **head of technical support** for the app. They are anxious about major changes to the app because they must update all of the help documentation about how to use it. They also need to instruct the tech support team about how to troubleshoot new problems. They have deep expertise in virtually every feature of the app. They believe major feature changes should be signaled to users well in advance so that tech support isn't overburdened by users who could solve the problem on their own.
Fill the boxes that follow with the information you'll need to craft a data story.

Fill the boxes that follow with the information you'll need to craft a data story.

The marketing team	
*What **problem** facing your audience does your insight help to solve?*	
*How can you rephrase your insight as an **outcome**, the logical consequence of the problem?*	
*What **actions** does your audience need to take to solve their problem, based on your insight?*	
*How might your audience best **measure** the success of their actions?*	

The head of technical support	
*What **problem** facing your audience does your insight help to solve?*	
*How can you rephrase your insight as an **outcome**, the logical consequence of the problem?*	
*What **actions** does your audience need to take to solve their problem, based on your insight?*	
*How might your audience best **measure** the success of their actions?*	

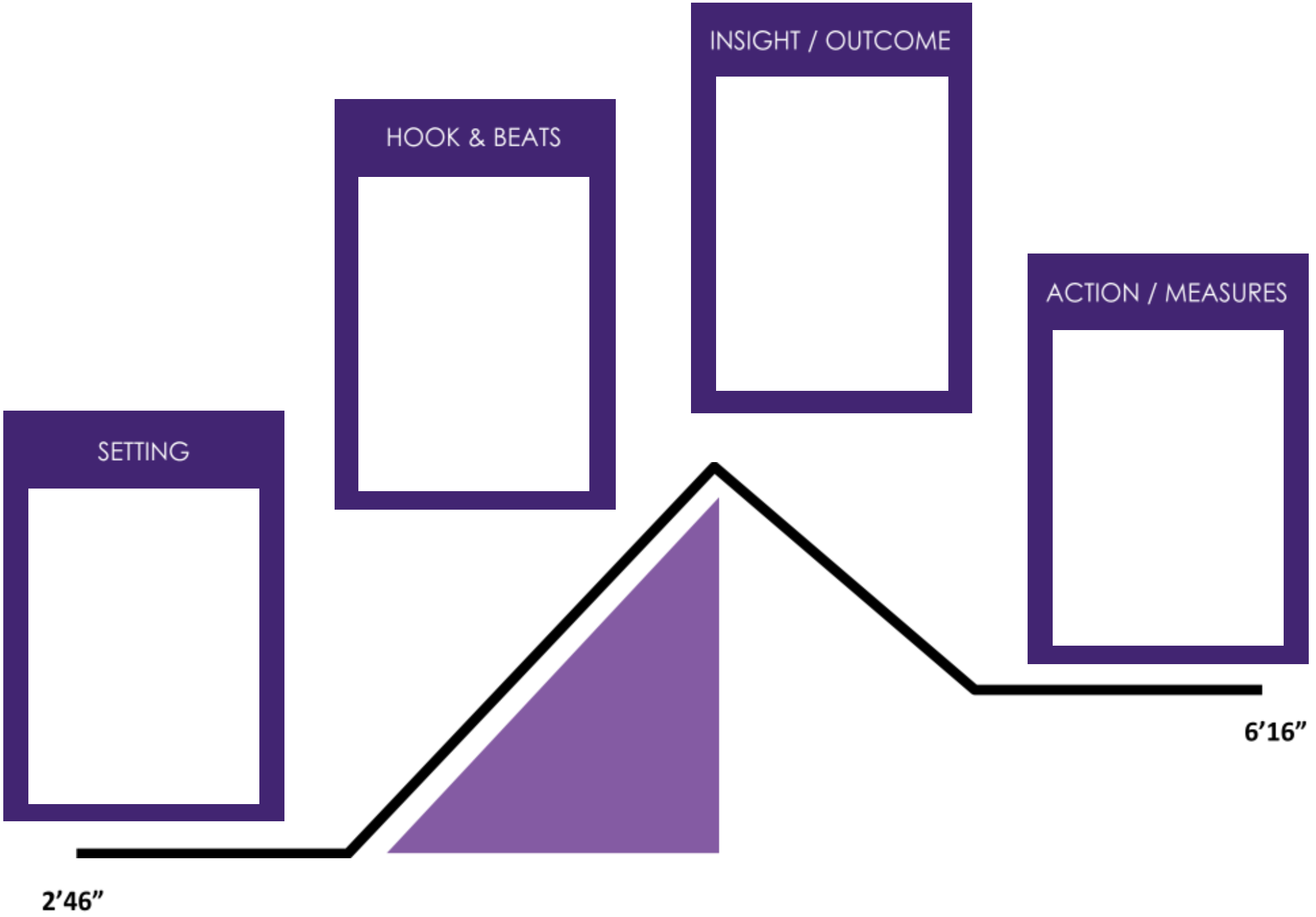
Activity: reverse storyboarding

A helpful exercise for getting more comfortable building your own data stories is to ***reverse engineer storyboards*** from other data stories.

Watch the segment of this video beginning at **2'46"** ("By visualizing this information we turned it into a landscape…") and ending at **6'16"** ("…and different patterns can be revealed"). It tells the story of a data visualization called "Mountains out of molehills."

Be aware of the different appeals to reason, emotion, and authority that the presenter uses. Note how the audience reacts to different parts of the story.

Now watch the video a second time. While watching, fill in the blanks on the diagram on the following page.


INSIGHT / OUTCOME
HOOK & BEATS
ACTION / MEASURES
SETTING
6’16”
2’46”

Additional resources

Tool selection checklist

- ☐ How much does the tool cost?
- ☐ Do I need approval from my organization to access, install, or use the tool?
- ☐ Is the tool relatively simple for new and/or non-technical users?
- ☐ How much data can the tool support?
- ☐ What types of visualizations can I create?
- ☐ Can I customize my visualizations?
- ☐ Can I easily move my data in and out of the tool?
- ☐ Can I host my data and visualizations online? Where?
- ☐ Is the tool actively maintained by its creators?
- ☐ Is the tool and my data protected against hackers and malware?
- ☐ Does the tool allow more than one person to work on the same project at the same time?
- ☐ Under the terms of service, is my information private or public?
- ☐ Does the tool work with the operating systems I need it to?
- ☐ Will the tool output properly on various devices and browsers?

Data visualization checklist

This checklist is based on the concepts covered during class. It is meant to be used as a guide for the development of high impact visualizations; however, the guidelines may be broken intentionally to make a point.

1. Overall Message

- ☐ The type of graph is appropriate for the data
- ☐ Graph highlights the significant findings or conclusions and gives the audience the appropriate takeaway message

2. Text

- ☐ Text size is hierarchical (e.g., titles are larger than labels, which are larger than axis labels, which are larger than source information)
- ☐ Text size is readable for the selected format (e.g., paper, screen)
- ☐ Labels have been placed directly next to the relevant data
- ☐ There are no extraneous or redundant data labels

3. Color

- ☐ The background is minimal and not distracting
- ☐ Visualization uses colors that are consistent with the rest of materials in which it is used
- ☐ Special effects have been removed or reduced (e.g., bolding, shading)
- ☐ Color is used effectively to make the audience focus on the most important pieces of information
- ☐ Color has been used consistently (i.e., changes in color are used to reinforce a change in topic or tone)
- ☐ Boldness, saturation, and/or brightness have been varied to distinguish colors for colorblind users
- ☐ Colors chosen are appropriate for the emotion you want to arouse in the audience
- ☐ If applicable, brand colors have been applied

4. Lines

- ☐ Visualization is free of unneeded borders
- ☐ Unnecessary gridlines have been deleted and necessary gridlines have been lightened
- ☐ Axes do not have unnecessary tick marks or axis lines

5. Size / Arrangement

- ☐ Proportions are depicted accurately
- ☐ Items of almost equal importance are sized similarly
- ☐ If there's one really important piece of information, it is BIG
- ☐ Bars/columns are ordered by descending or ascending value (unless trying to show trend over time)
- ☐ In dashboards, charts that are closer to one another actually more related than other charts that are not as close together
- ☐ In dashboards, the most important pieces of information are placed at the top of the page
- ☐ In dashboards, elements are placed in a way that feels natural

Where to learn more

- Books by Edward Tufte
- Publications by Colin Ware
- Books by David McCandless, as well as his updated version of "Mountains out of molehills"
- TED Talks from Hans Rosling
- *DataStory* by Nancy Duarte
- *Effective Data Storytelling* by Brent Dykes
- *Follow up on Bem's PSI research* by Steven Novella
- *Made to Stick: Why Some Ideas Survive and Others Die* by Chip and Hank Heath
- *Storytelling with Data* by Cole Nussbaumer Knaflic (as well the rest of her website)
- *The Building Blocks of Visual Design* by Teo Yu Siang, for the Interaction Design Foundation
- *"When Numbers Mislead"* by Stephanie Coontz, in *The New York Times*
- *Why the Journal of Personality and Social Psychology Should Retract Article DOI: 10.1037/a0021524 "Feeling the Future: Experimental evidence for anomalous retroactive influences on cognition and affect" by Daryl J. Bem* by Ulrich Schimmack