

# Before We Begin

## Pre-workshop notes



Use headphones



Everyone their own laptop



Quiet space



Reliable internet connection



Camera on

## online etiquette



Mute yourself



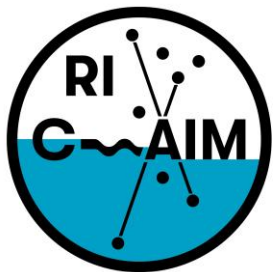
Press spacebar to unmute



Want to speak



Yes / agree



RHODE ISLAND CONSORTIUM FOR  
Coastal Ecology  
Assessment  
Innovation &  
Modeling



This material is based upon work supported in part by the National Science Foundation under EPSCoR Cooperative Agreement #OIA-1655221

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.



# Data Visualization Best Practices

Day 1: Best Practices and useful tools

# Agenda

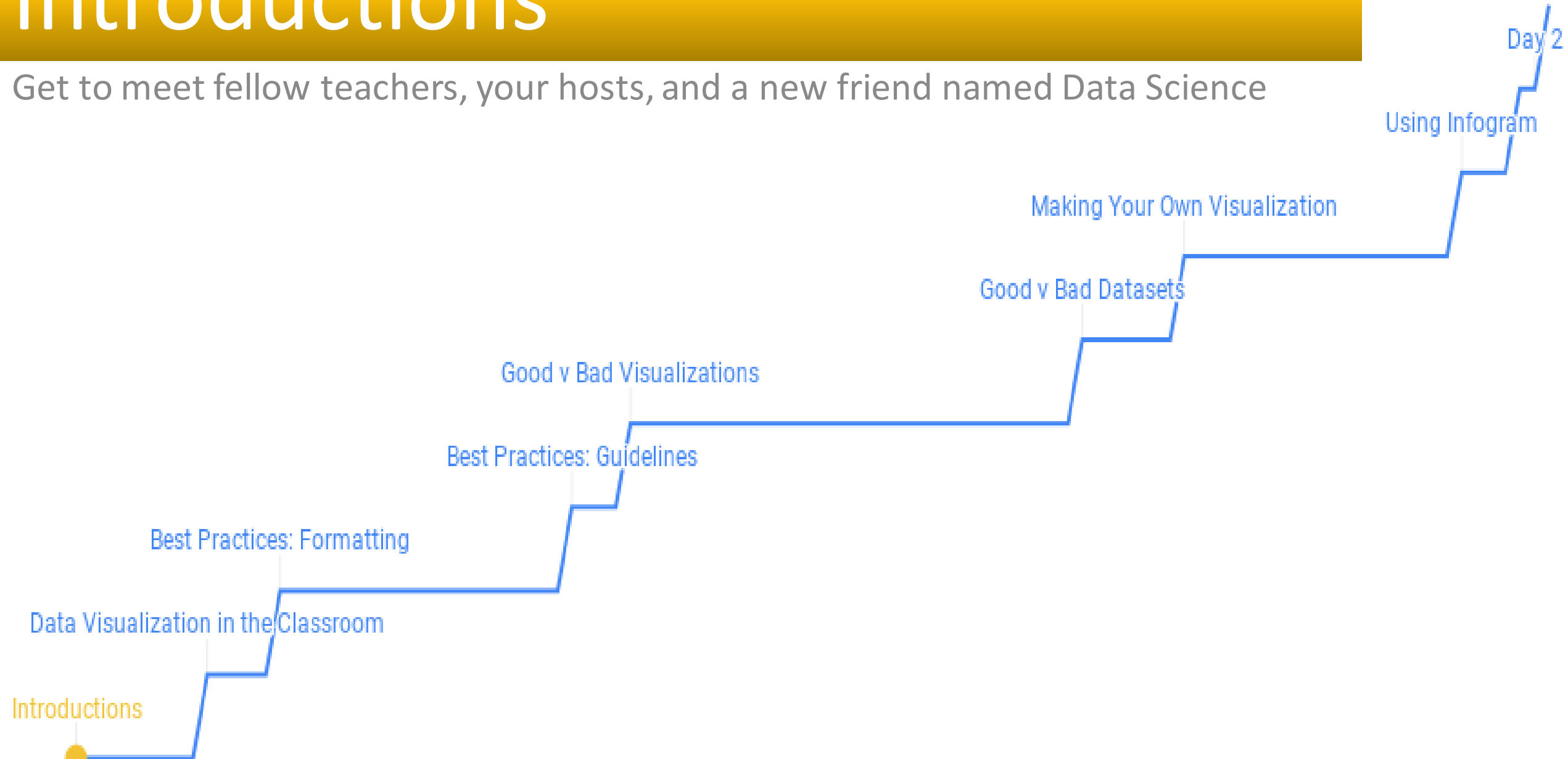
- 9:00 – 9:20 Introductions: Who are we?
  - Prof. Maia welcome notes
  - Prof. Hamouda welcome notes
  - Student introductions
- 9:20 – 9:25 Welcome from Dr. Anna Cano Morales, Associate Vice President for Community, Equity and Diversity (Video)
- 9:25- 9:45 Welcome from Joe Devin: Tech Collective Executive Director (Video)
- 9:45 – 10:00 Teachers' introductions
- 10:00 – 10:10 Break
- 10:10 – 10:55 Introductions to Good vs. Bad Visualizations
- 10:55 – 11:05 Break
- 11:05 – 11:30 Good vs. Bad Visualizations
- 11:30 – 12:00 Data Visualization tools
- 12:00 – 12:30 Q&A/discussion, Planning for Day 2

# Agenda

- **9:00 – 9:20 Introductions: Who are we?**
  - Prof. Maia welcome notes
  - Prof. Hamouda welcome notes
  - Student introductions
- 9:20 – 9:25 Welcome from Dr. Anna Cano Morales, Associate Vice President for Community, Equity and Diversity (Video)
- 9:25- 9:45 Welcome from Joe Devin: Tech Collective Executive Director (Video)
- 9:45 – 10:00 Teachers' introductions
- 10:00 – 10:10 Break
- 10:10 – 10:55 Introductions to Good vs. Bad Visualizations
- 10:55 – 11:05 Break
- 11:05 – 11:30 Good vs. Bad Visualizations
- 11:30 – 12:00 Data Visualization tools
- 12:00 – 12:30 Q&A/discussion, Planning for Day 2

# Introductions

Get to meet fellow teachers, your hosts, and a new friend named Data Science



# Who are we?



Names, schools, what we do



What have we been working on?



What is NSF and RI C-AIM?

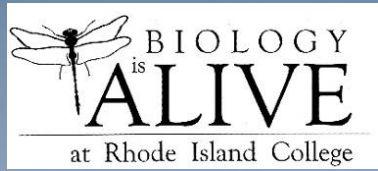
Who are we?

RIC PI

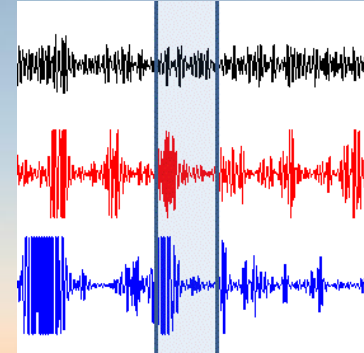
Dr. Anabela Da Maia



Dr. Maia



Dr. Anabela Maia, PhD  
Department of Biology  
Rhode Island College

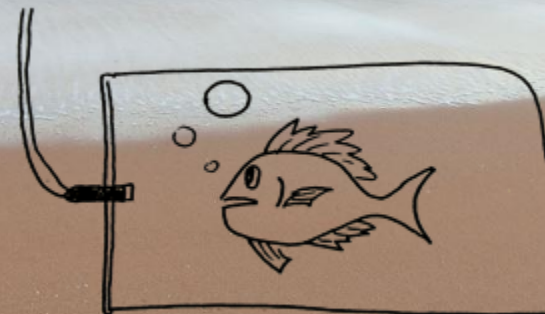


[aresendedamaia@ric.edu](mailto:aresendedamaia@ric.edu)



[@AnabelaM](https://twitter.com/AnabelaM)

<http://maialabric.wordpress.com/>





# Personal Narrative



RHODE  
ISLAND  
COLLEGE



# RI C-AIM Who we are?



RHODE ISLAND COLLEGE









 

## The Rhode Island Consortium for Coastal Ecology Assessment, Innovation and Modeling

RI NSF EPSCoR Award #OIA-1655221



@RIEPSCoR

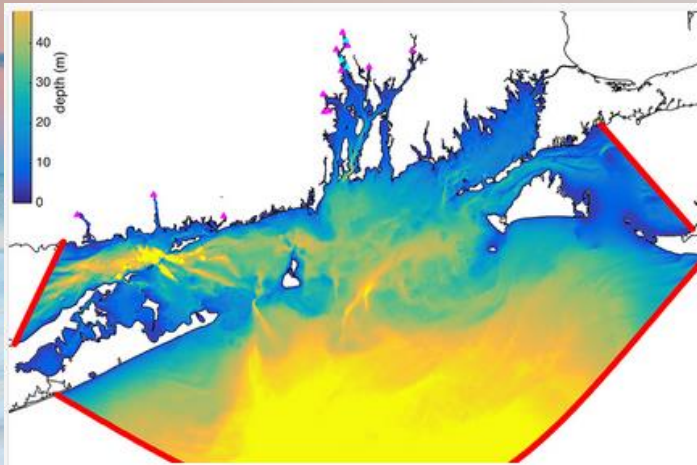
 <p><b>THE UNIVERSITY OF RHODE ISLAND</b></p> <p>University of Rhode Island RI C-AIM lead institution</p> <p>VISIT →</p>	 <p><b>BROWN</b></p> <p>Brown University data management, ocean modeling</p> <p>VISIT →</p>	 <p><b>BRYANT UNIVERSITY</b></p> <p>Bryant University marine food webs, microbial dynamics</p> <p>VISIT →</p>	 <p><b>PROVIDENCE COLLEGE</b></p> <p>Providence College species identification, visualization</p> <p>VISIT →</p>
 <p><b>RHODE ISLAND COLLEGE</b></p> <p>Rhode Island College microbiological modeling, workforce development</p> <p>VISIT →</p>	 <p><b>RHODE ISLAND SCHOOL OF DESIGN</b></p> <p>Rhode Island School of Design data visualization, graphical research translation</p> <p>VISIT →</p>	 <p><b>Roger Williams University</b></p> <p>Roger Williams University microplastics detection, undergraduate research training</p> <p>VISIT →</p>	 <p><b>SALVE REGINA UNIVERSITY</b></p> <p>Salve Regina University microfluidic systems, planktonic diversity characterization</p> <p>VISIT →</p>

# RI C-AIM Research Thrusts



RHODE  
ISLAND  
COLLEGE

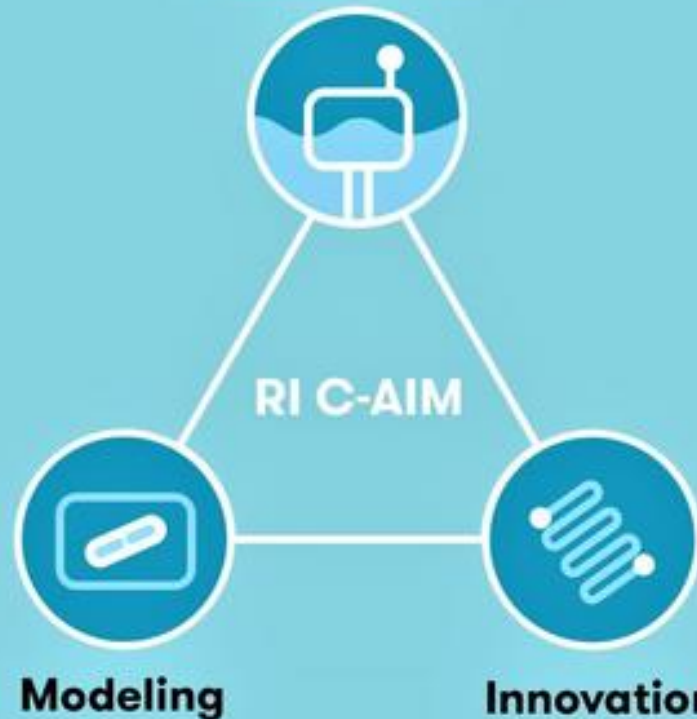
**Assessing Biological and Ecosystem Impacts (RT 1)** Through the creation of the integrated Bay Observatory, researchers are gathering real-time ecological data from Narragansett Bay and surrounding watersheds, providing highly-detailed assessments of the bay's ecosystems and interactions between nutrients and living species.



## **Predicting Ecosystem Response Through Integration (RT 2)**

Combining present and historical data, scientists are creating models of the bay's ecosystems, from nutrient distribution to changes in species habits. Models will also develop a clearer picture of how human behavior affects, and is affected by, coastal environments.

## **Assessment RT1**



RT2

RT3



## **Enabling Technologies for Improved Detection (RT 3)**

Researchers are developing new platforms to detect chemical and biological interactions in Narragansett Bay, from paper-based devices to biosensor systems. This technology will more accurately detect the changing environments of Narragansett Bay, and give the public a chance to collect data.



RHODE ISLAND COLLEGE

# RI C-AIM Inter Thrusts



**Educational initiatives** are incorporating RI C-AIM research into college-level courses and senior capstone projects



**SURF & SURF+** are paid programs that provide undergraduates with first-hand experience in research

## IT3 – Stakeholder engagement



**Visualization and Imaging (IT 1)** Across thrusts, RI C-AIM is developing novel approaches to visualization of scientific observations in complex ecosystems through collaborations between engineers, designers, artists and oceanographers, thus fostering greater understanding from industry leaders, policy makers and the public.

## IT2 – Increasing representation in STEAM



**Diversity Action Committee (DAC):** Resources for students and faculty to achieve inclusive collaborations



**Career Development:** RI C-AIM offers many opportunities in research and professional skills training

## IT1 – Visualization and Imaging

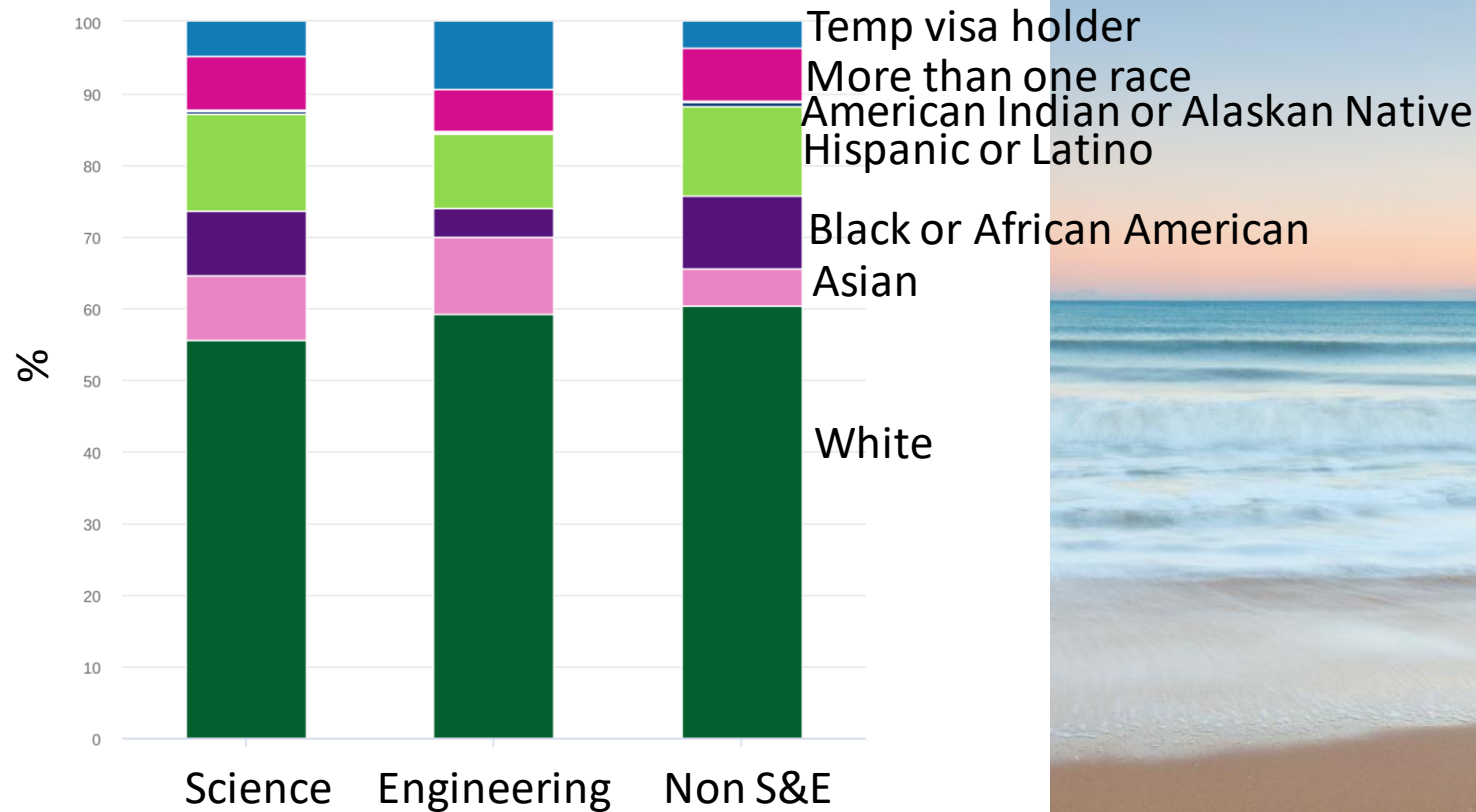


Scientists and graphic artists are coming together to develop media which help explain research in stimulating and novel ways.

# Empowering and Recruiting URM



Bachelor's degrees earned by ethnicity, race and citizenship in 2016 (source NSF)

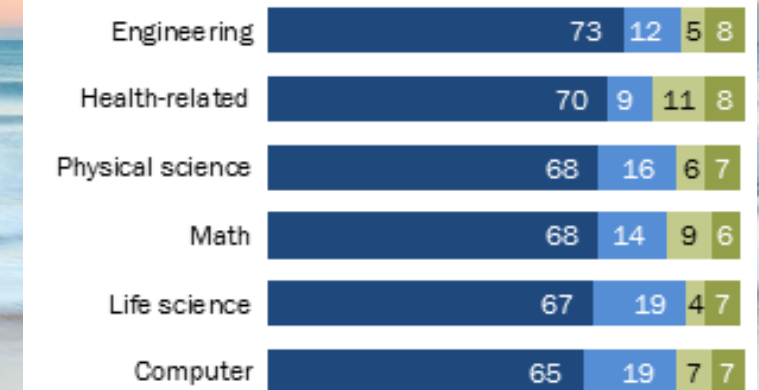


## Blacks and Hispanics underrepresented across most STEM job clusters

% of employed in each occupational group who are ...



Among those who work in \_\_\_ jobs



Note: Based on employed adults ages 25 and older. Whites, blacks and Asians include only non-Hispanics. Hispanics are of any race. Other and mixed race non-Hispanics are not shown. Engineering includes architects. STEM stands for science, technology, engineering and math.

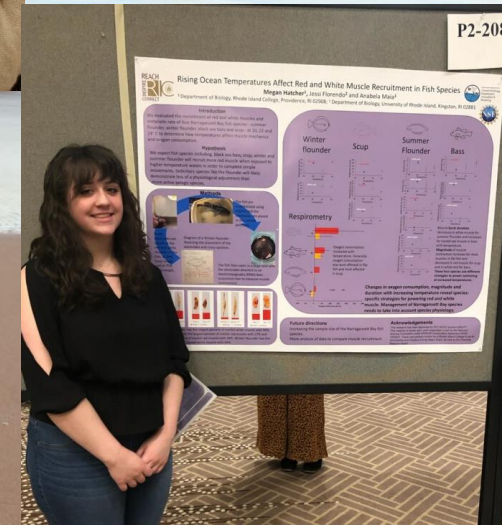
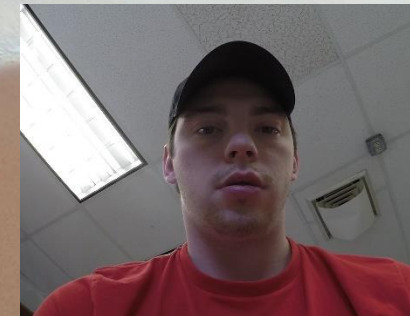
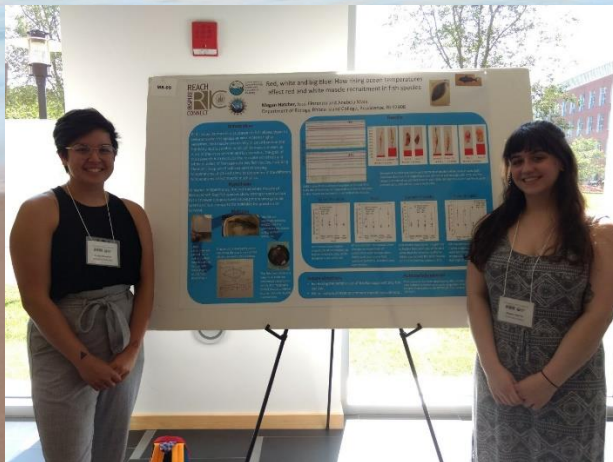
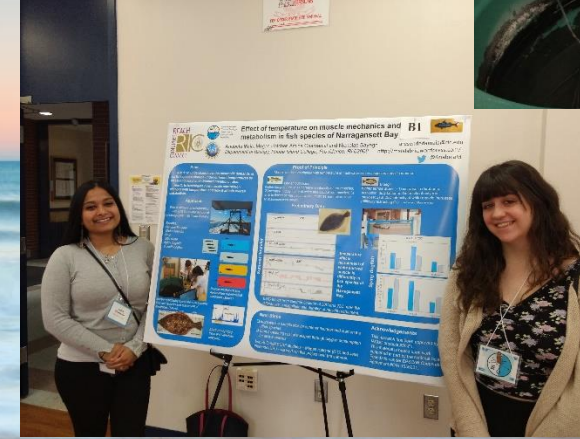
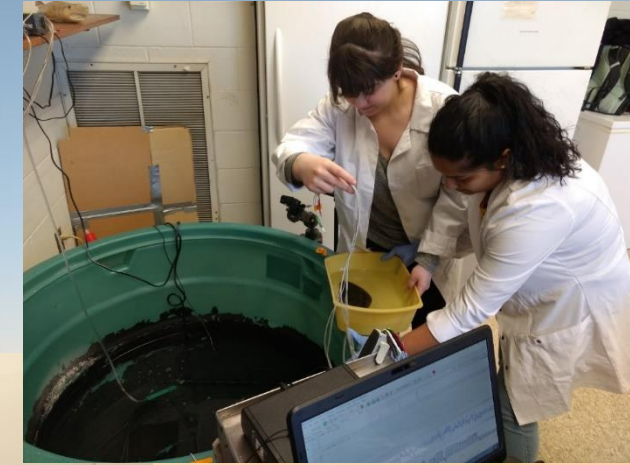
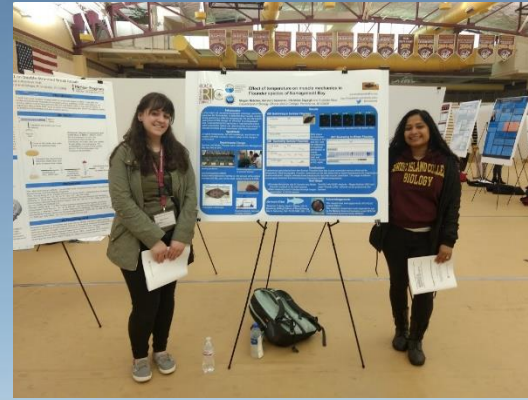
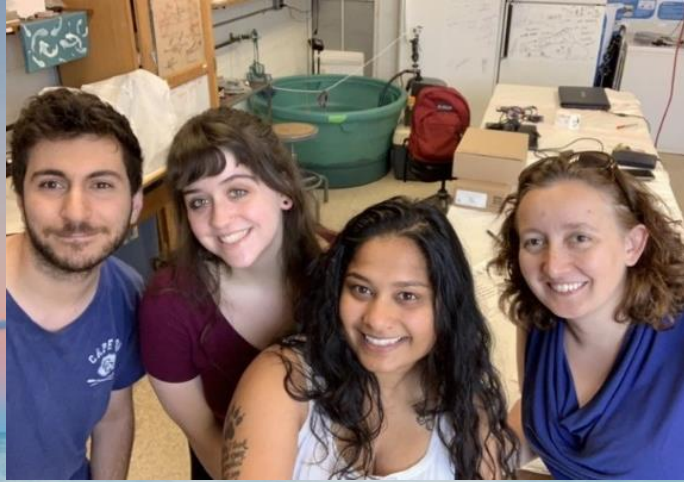
Source: Pew Research Center analysis of 2014-2016 American Community Survey (IPUMS).

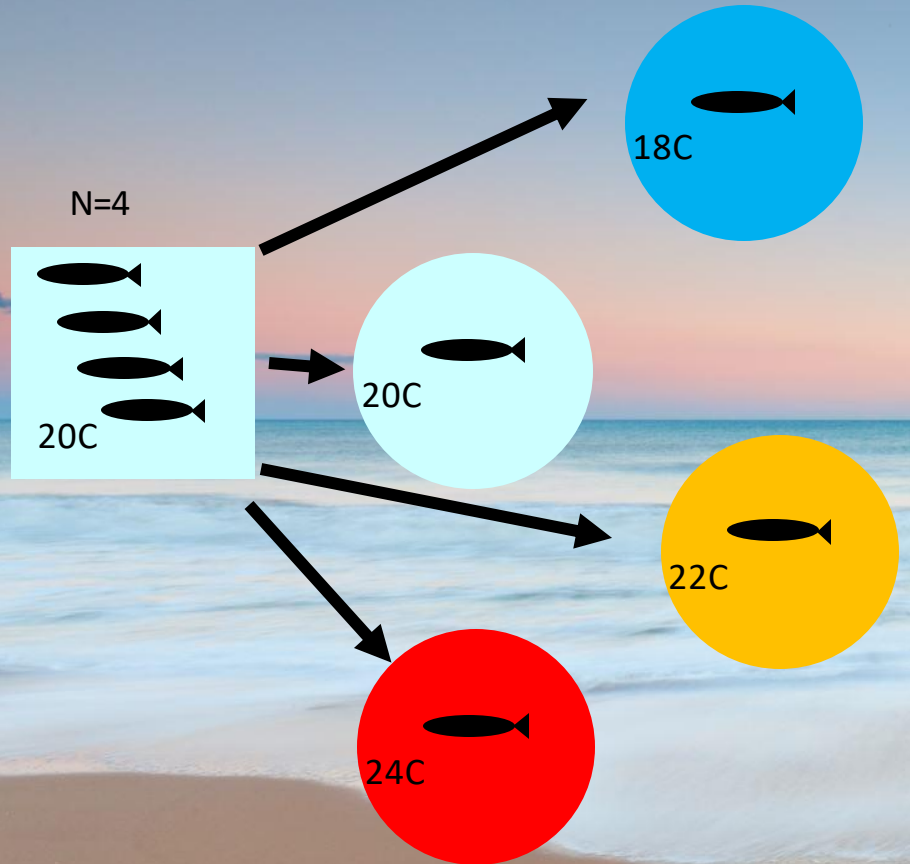
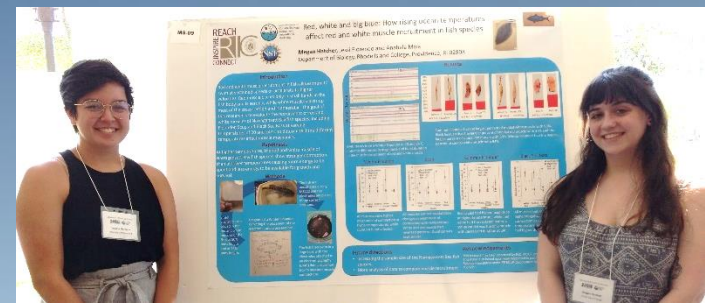
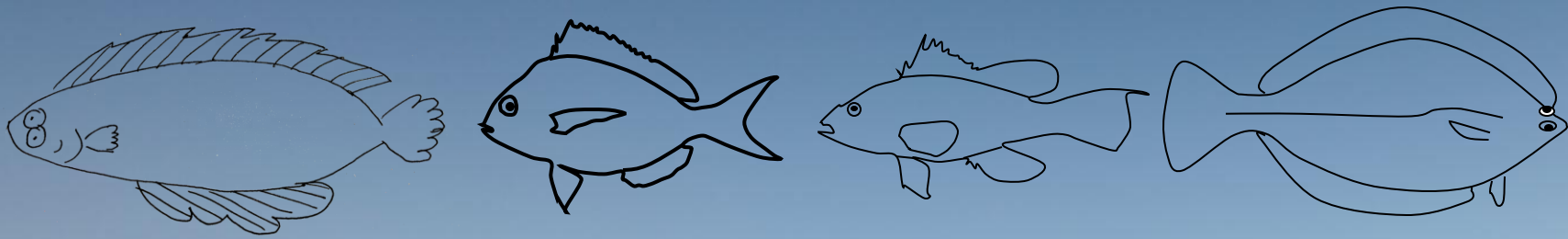
"Women and Men in STEM Often at Odds Over Workplace Equity"

# RI C-AIM at RIC



RHODE ISLAND COLLEGE






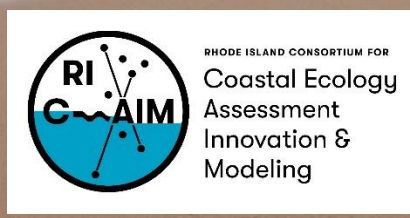
How are fish species of Narragansett Bay affected by rising summer temperatures?



Are muscle mechanics and oxygen consumption limiting factors?



 @AnabelaM  
aresendedamaia@ric.edu

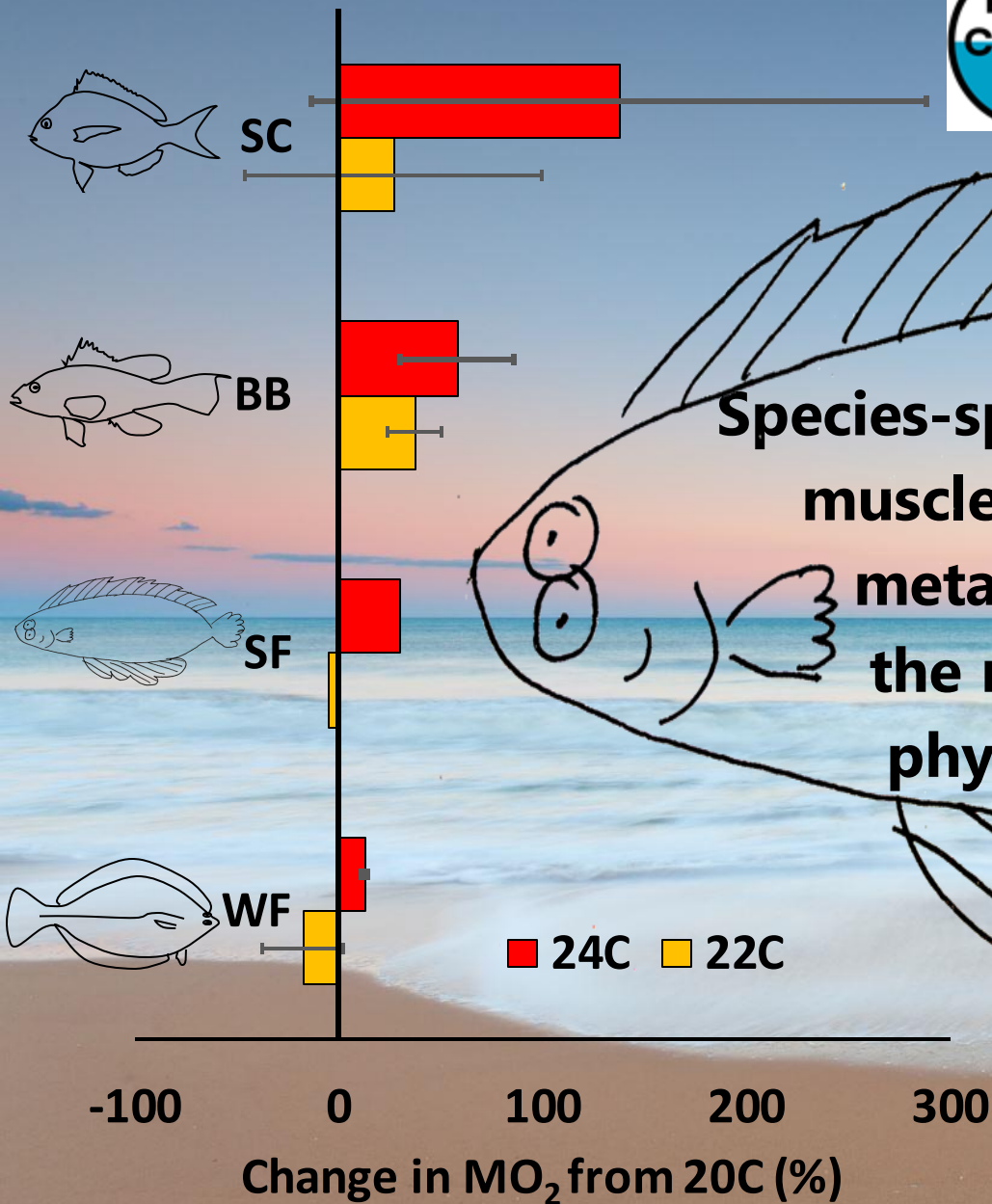
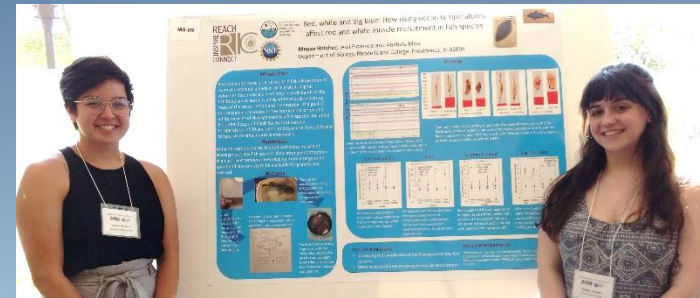


RHODE ISLAND COLLEGE

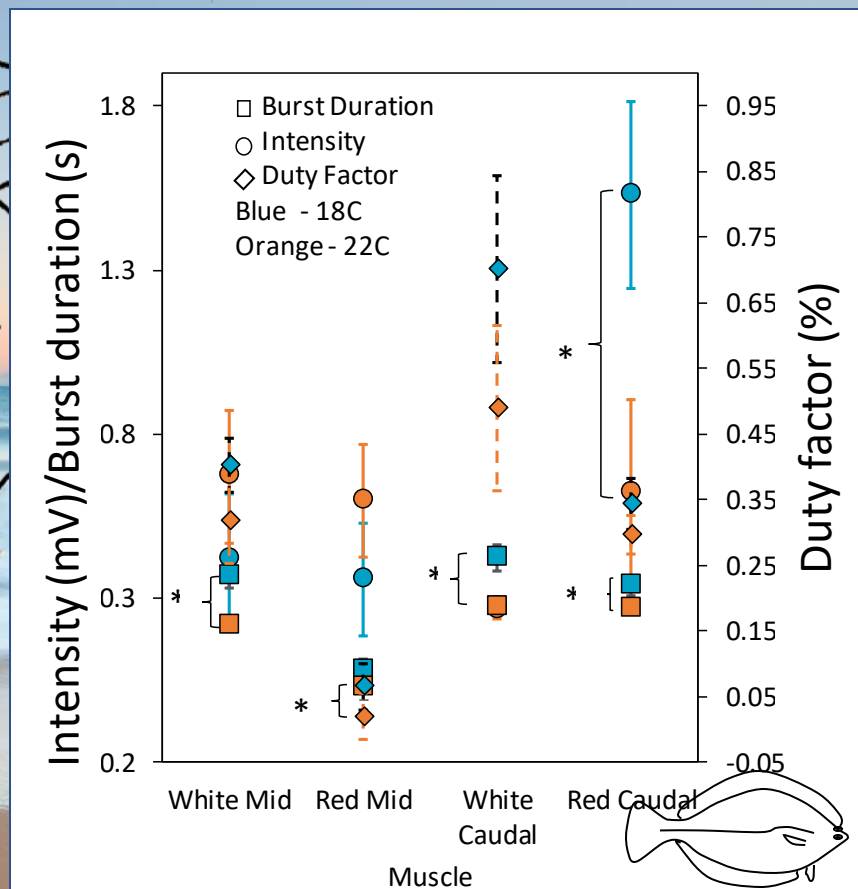
<http://maialabric.wordpress.com>



RHODE ISLAND COLLEGE



**Species-specific changes in muscle mechanics and metabolic rate highlight the need for more physiological studies.**



Higher dependency on red muscles at higher temperatures.

Oxygen consumption is less affected in flatfish species.

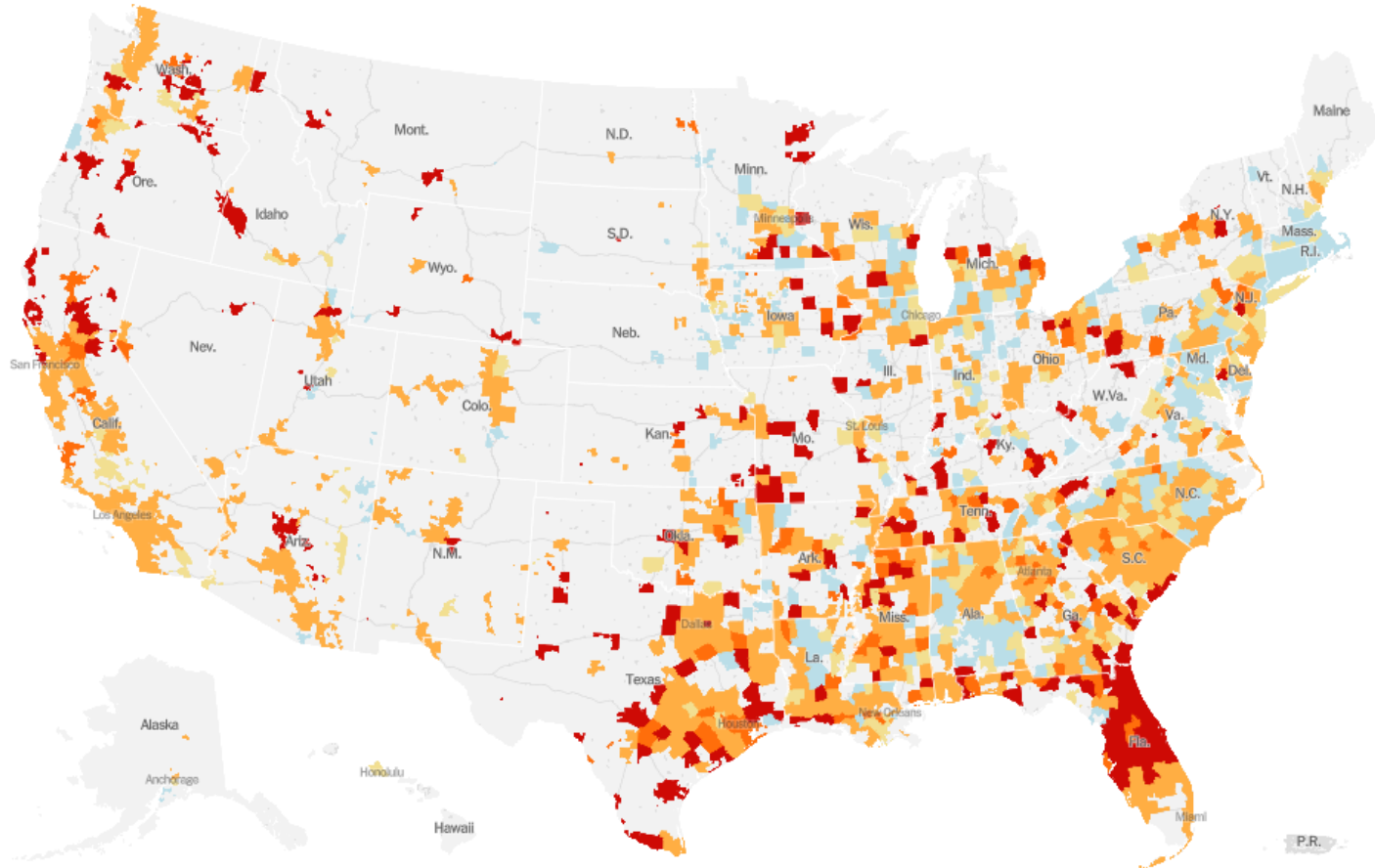


# Data Visualization in Biology

## COVID-19



RHODE ISLAND COLLEGE



mapbox

Sources: State and local health agencies and hospitals.

[About this data](#)

### The New York Times

Hot spots	Total cases	Deaths	Per capita
How the number of new cases has changed in the last two weeks			
Falling	About the same	Rising →	Few or no cases

Double-click to zoom into the map.

# Data Visualization in Biology

## COVID-19

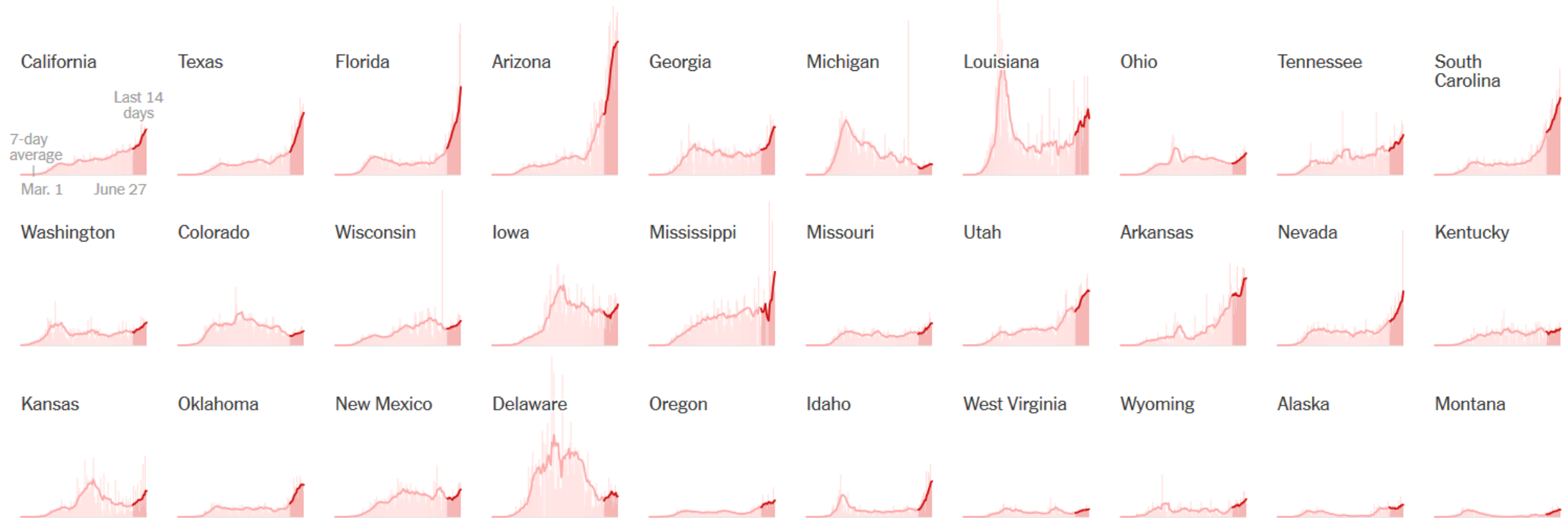


### Where new cases are increasing

Cases per capita

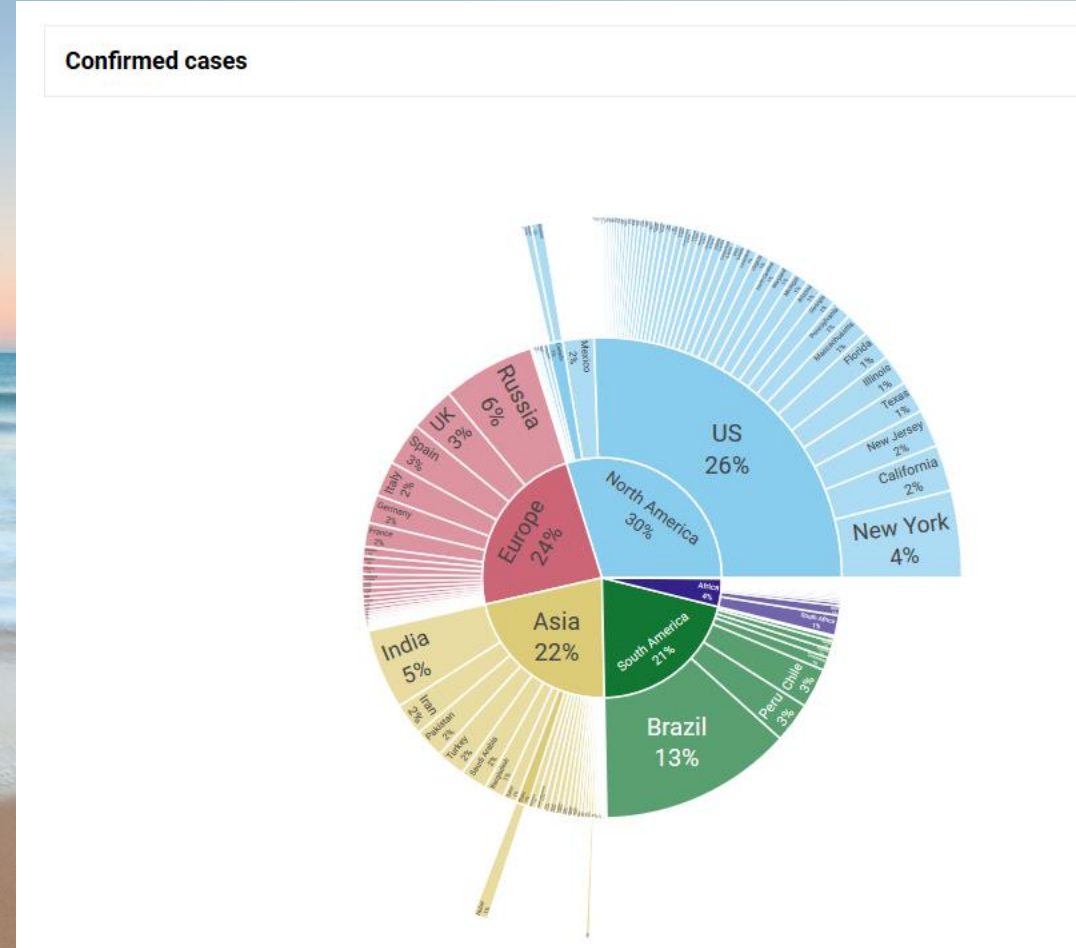
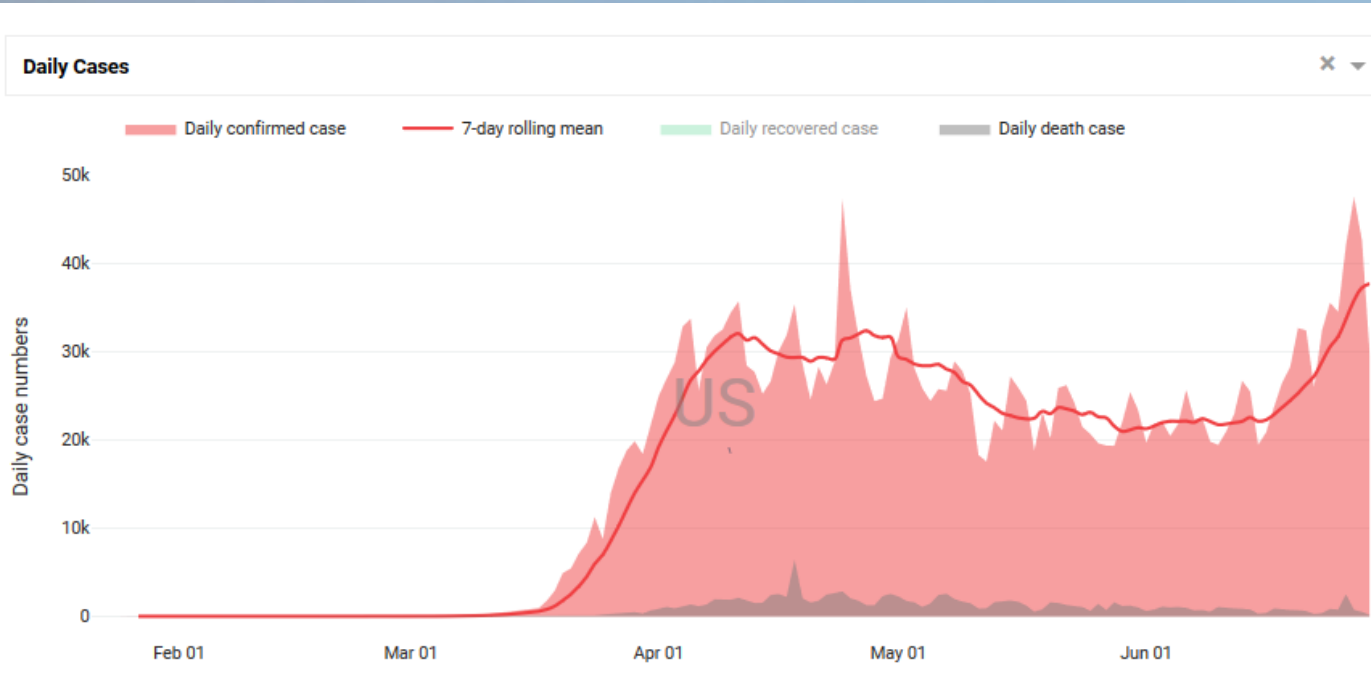
Total cases

All charts are shown on the same scale. Click a state to see detailed map page.



# Data Visualization in Biology

## COVID-19



Source: <https://dash-coronavirus-2020.herokuapp.com/>

# Data Visualization in Biology

## Epilepsy



RHODE  
ISLAND  
COLLEGE

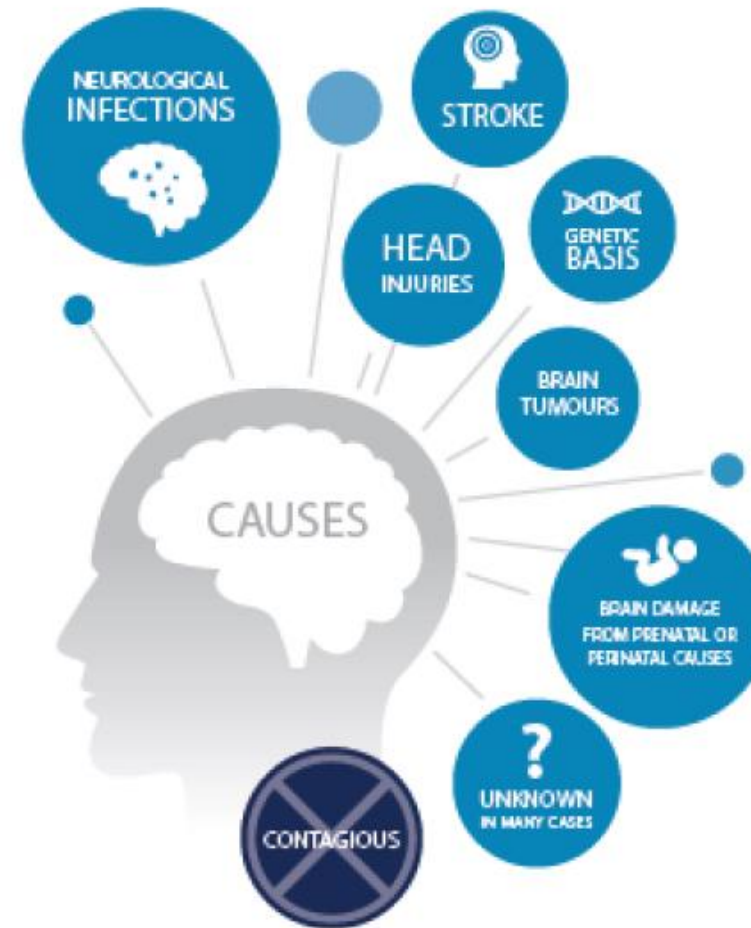
## WHAT IS epilepsy?

A NEUROLOGICAL CONDITION characterized by *recurrent seizures*

Seizures are due to *brief disturbances* in the *electrical functions* of the brain



Epilepsy affects people of all ages



World Health  
Organization

# Data Visualization in Biology

## Epilepsy



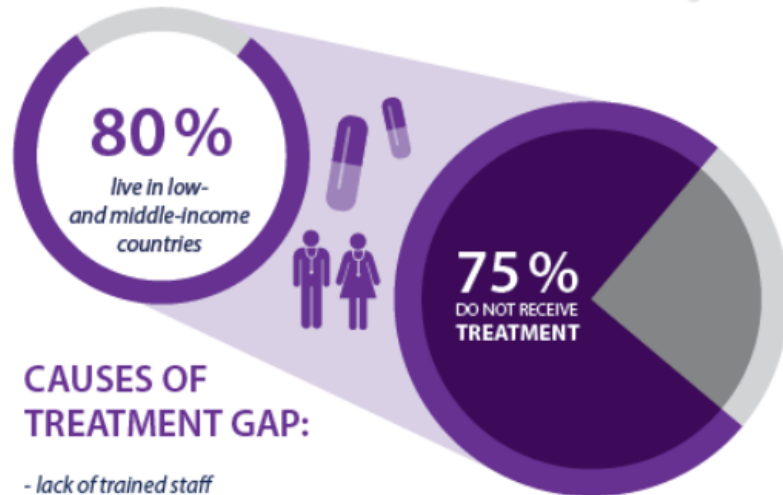
RHODE ISLAND COLLEGE

### What is the **IMPACT** of epilepsy?

**50 000 000**

More than 50 million people are living with epilepsy globally

**3-6** TIMES  
GREATER  
**RISK**  
OF PREMATURE  
DEATH



#### CAUSES OF TREATMENT GAP:

- lack of trained staff
- poor access to anti-epileptic medicines
- societal misconceptions
- poverty
- low prioritization for the treatment of epilepsy



### What is the **SOLUTION**?

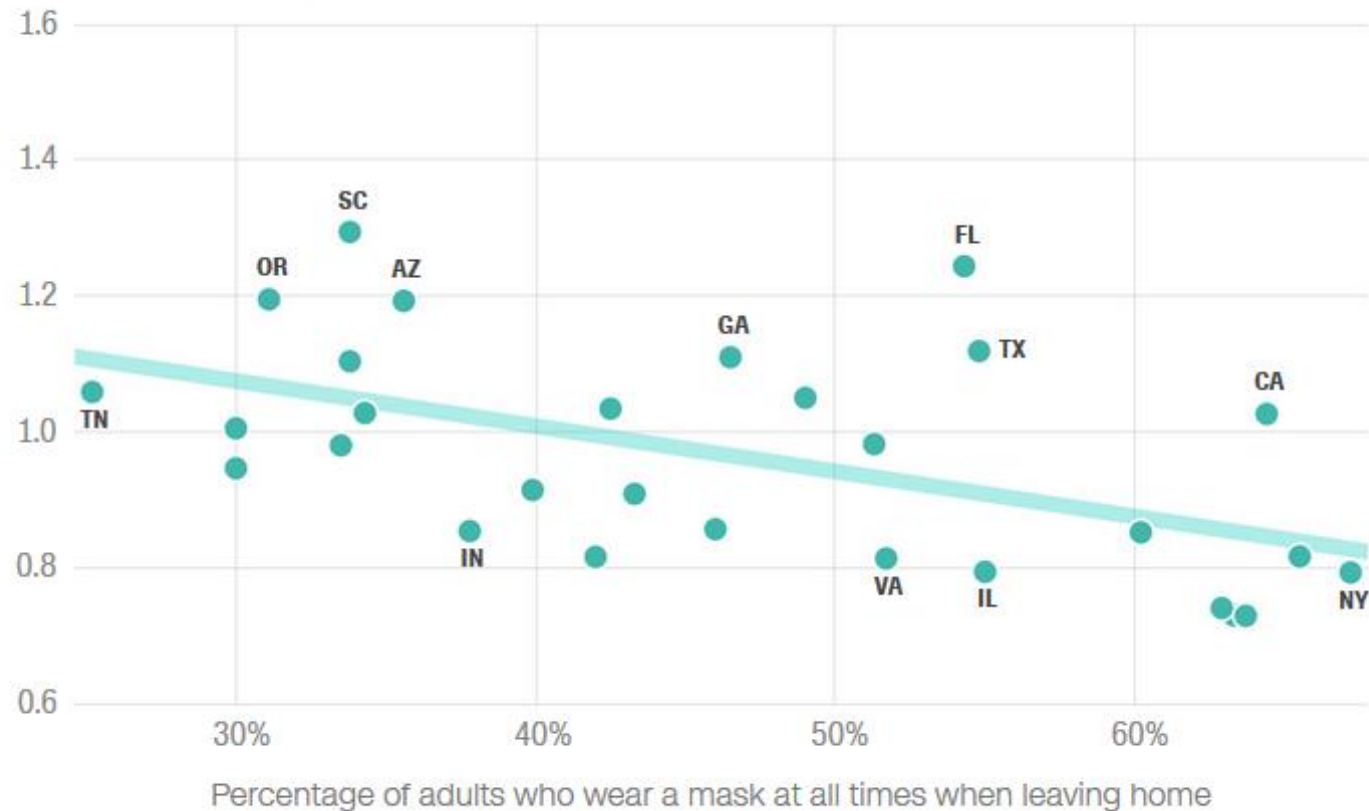




RHODE ISLAND COLLEGE

# Data are Powerful!

Rate of Covid-19 reproduction, May 8 - June 22



Note: This chart includes the 28 states for which there was a sufficient sample size to break out separately. Sample sizes for each of those states range from 75 to 764 survey respondents.

Source: Axios/Ipsos and rt.live

Graphic: Harry Enten and Janie Boschma, CNN

# Questions?



Native Scientist "Around the World with Portuguese Scientists"

Who are we?  
RIC Co-PI  
Dr. Sally Hamouda

Research  
Interests

- CS EDUCATION
- DATA MINING
- LEARNING ANALYTICS

WORKED AS

- EDUCATOR
- RESEARCHER
- MENTOR

CONTACT

- SHAMOUDA@RIC.EDU



# Who are we? Our Students



Lauren Cenedella



Meghan Vanschalkwyk



Destiny Gonzalez



Matthew Spaulding



Samantha Palacio

# Agenda

- 9:00 – 9:20 Introductions: Who are we?
  - Prof. Maia welcome notes
  - Prof. Hamouda welcome notes
  - Student introductions
- **9:20 – 9:25 Welcome from Dr. Anna Cano Morales, Associate Vice President for Community, Equity and Diversity (Video)**
- 9:25- 9:45 Welcome from Joe Devin: Tech Collective Executive Director (Video)
- 9:45 – 10:00 Teachers' introductions
- 10:00 – 10:10 Break
- 10:10 – 10:55 Introductions to Good vs. Bad Visualizations
- 10:55 – 11:05 Break
- 11:05 – 11:30 Good vs. Bad Visualizations
- 11:30 – 12:00 Data Visualization tools
- 12:00 – 12:30 Q&A/discussion, Planning for Day 2

Dr. Anna Cano Morales, As  
sociate Vice President for  
Community, Equity and  
Diversity



# Agenda

- 9:00 – 9:20 Introductions: Who are we?
  - Prof. Maia welcome notes
  - Prof. Hamouda welcome notes
  - Student introductions
- 9:20 – 9:25 Welcome from Dr. Anna Cano Morales, Associate Vice President for Community, Equity and Diversity (Video)
- **9:25- 9:45 Welcome from Joe Devin: Tech Collective Executive Director (Video)**
- 9:45 – 10:00 Teachers' introductions
- 10:00 – 10:10 Break
- 10:10 – 10:55 Introductions to Good vs. Bad Visualizations
- 10:55 – 11:05 Break
- 11:05 – 11:30 Good vs. Bad Visualizations
- 11:30 – 12:00 Data Visualization tools
- 12:00 – 12:30 Q&A/discussion, Planning for Day 2

Joe Devin: Tech Collective  
Executive Director

## Data Analysts + Business Intelligence

Responsible for analyzing data from multiple disparate sources to discover previously hidden insight, determine meaning behind the data, and make business-specific recommendations.

Visual reporting to effectively communicate findings



# Agenda

- 9:00 – 9:20 Introductions: Who are we?
  - Prof. Maia welcome notes
  - Prof. Hamouda welcome notes
  - Student introductions
- 9:20 – 9:30 Welcome from Joe Devin: Tech Collective Executive Director (Video)
- 9:30 – 9:45 Welcome from Dr. Anna Cano Morales, Associate Vice President for Community, Equity and Diversity (Video)
- **9:45 – 10:00 Teachers' introductions**
- 10:00 – 10:10 Break
- 10:10 – 10:55 Introductions to Good vs. Bad Visualizations
- 10:55 – 11:05 Break
- 11:05 – 11:30 Good vs. Bad Visualizations
- 11:30 – 12:00 Data Visualization tools
- 12:00 – 12:30 Q&A/discussion, Planning for Day 2

# Who are you?



Name, school, what you teach



Why you were interested in this particular workshop?



What do you hope to learn?



What experience with data visualization or data science do you have?

# 10-minute break

## Pre-workshop notes



Use headphones



Everyone their own laptop



Quiet space



Reliable internet connection



Camera on

## online etiquette



Mute yourself



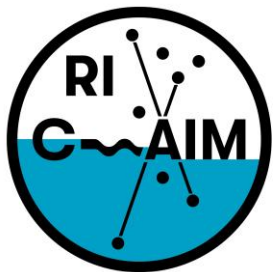
Press spacebar to unmute



Want to speak



Yes / agree



RHODE ISLAND CONSORTIUM FOR  
Coastal Ecology  
Assessment  
Innovation &  
Modeling



This material is based upon work supported in part by the National Science Foundation under EPSCoR Cooperative Agreement #OIA-1655221

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.



# Agenda

- 9:00 – 9:20 Introductions: Who are we?
  - Prof. Maia welcome notes
  - Prof. Hamouda welcome notes
  - Student introductions
- 9:20 – 9:30 Welcome from Joe Devin: Tech Collective Executive Director (Video)
- 9:30 – 9:45 Welcome from Dr. Anna Cano Morales, Associate Vice President for Community, Equity and Diversity (Video)
- 9:45 – 10:00 Teachers' introductions
- 10:00 – 10:10 Break
- **10:10 – 10:55 Introductions to Good vs. Bad Visualizations**
- 10:55 – 11:05 Break
- 11:05 – 11:30 Good vs. Bad Visualizations
- 11:30 – 12:00 Data Visualization tools
- 12:00 – 12:30 Q&A/discussion, Planning for Day 2

# What are they?

## **Data Science**

- Collecting, cleaning, and communicating data
- Use programming languages, have good people skills, be able to generate insights from the data

## **Data Visualization**

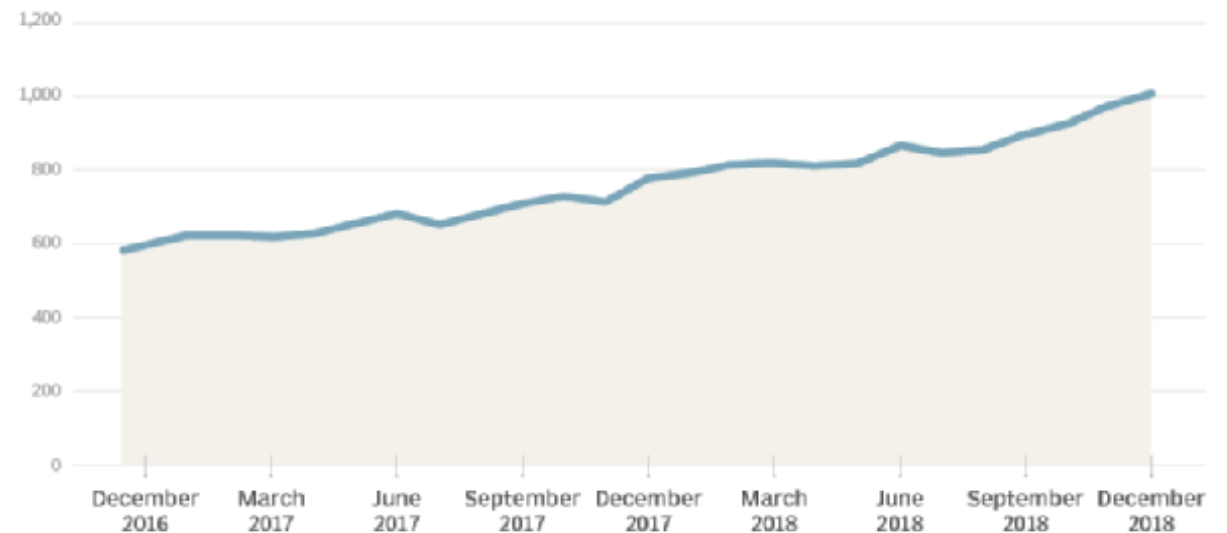
- Process of graphically presenting data to reveal its patterns, trends, and meaning
- Usually made for big data
- Can be reused with other data sets

# Data Scientists are in demand

Citing reports by LinkedIn, TechTarget explains that demand for data scientists outstrips the number of people that have the skills to fill those openings

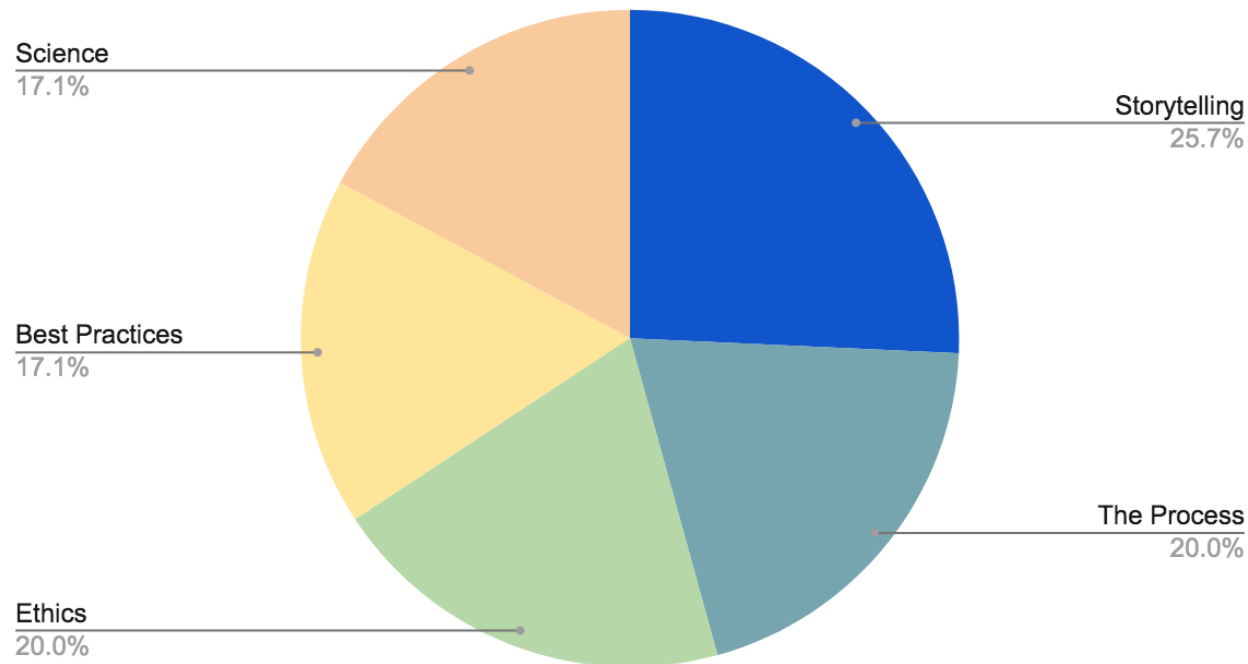
## Data scientists are in high demand

Data scientist job postings, per 1 million postings on Indeed



# What are people saying about data visualization?

Frequency of Concepts



The most common themes found in our research. All of them have almost the same importance.

- Storytelling: how you communicate the data
- The Process: all the steps necessary to analyze and act on data
- Best Practices: general guidelines on how to make a visualization
- Science: using the scientific method to answer a hypothesis/question through data
- Ethics: presenting data truthfully to the best of your ability

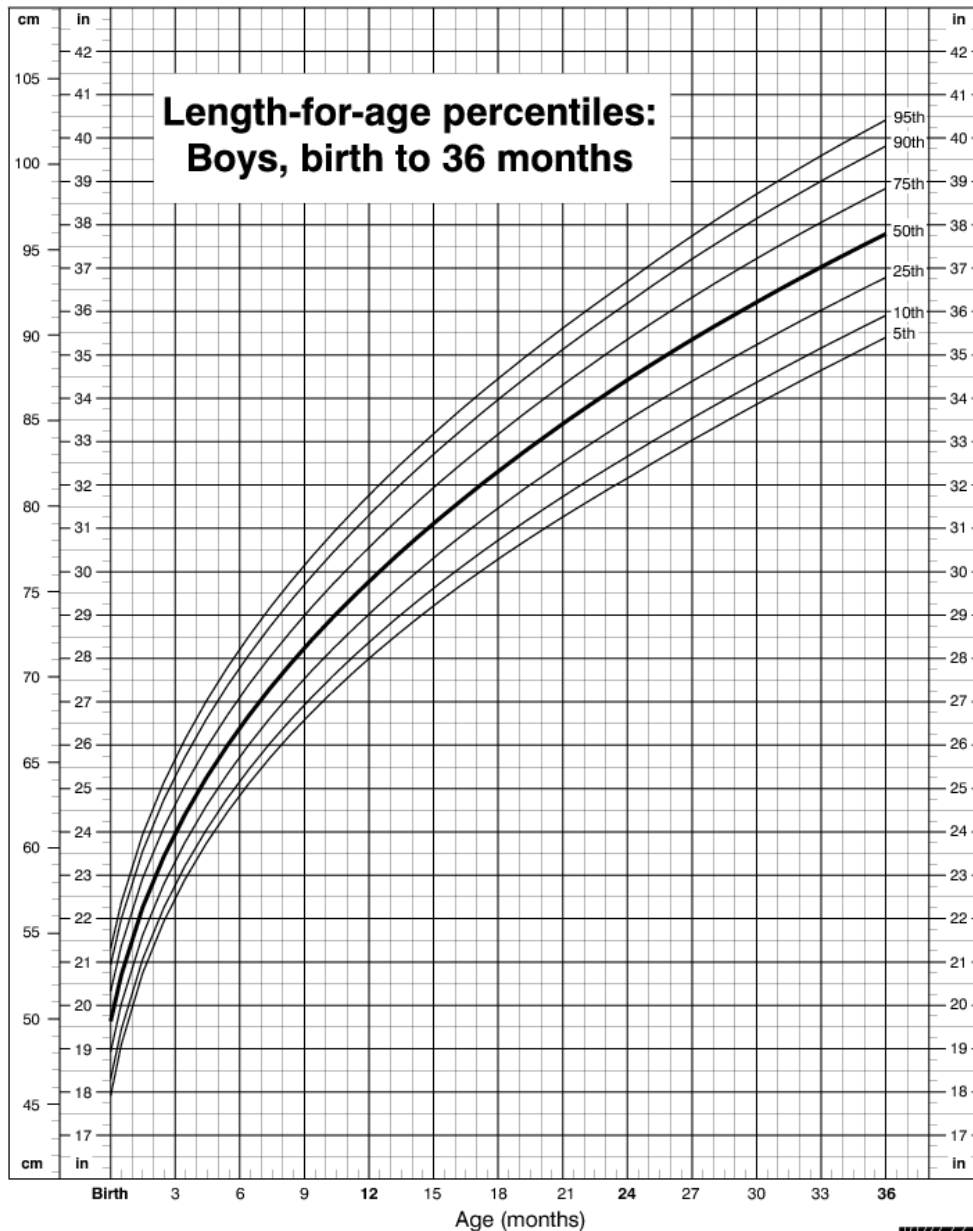
# A special note on ethics

ALWAYS ALWAYS ALWAYS question what you're looking at! Any visualization can be misleading or misinterpreted. Ask yourself:

- Where does this data come from?
- How was it collected? Is the sample size and diversity satisfactory?
- Is this chart an appropriate representation, or would a different style be better?
- What is this visualization trying to tell me?

**In the chapters of his book, he discusses the many ways a chart might lie. It might be a product of poor design or use incorrect data. It might include an inappropriate amount of data, conceal data, or not clearly communicate uncertainty. Or, it might suggest misleading patterns or pander to pre-existing prejudices.**

Bisoux, Tricia. "The Ethics of Data Visualization"



# A special note on ethics continued

- The height growth charts are one example of unethical visualizations – it shows an inappropriate amount of data (too little).
- The line chart implies continuous data, but Michelle Lampl found that infants grow in discontinuous stages when you look at them day by day (not by months). There could be no growth for weeks, and then all of a sudden they grow about 2 cm in one night. The graph should be more step-wise.



Michelle  
Lampl



# Data Visualization in the Classroom





# How does it apply to the classroom?



## Conventional ways

Math

Science

Computer Science



## Unconventional ways

Health

English

Art

Creating your syllabus

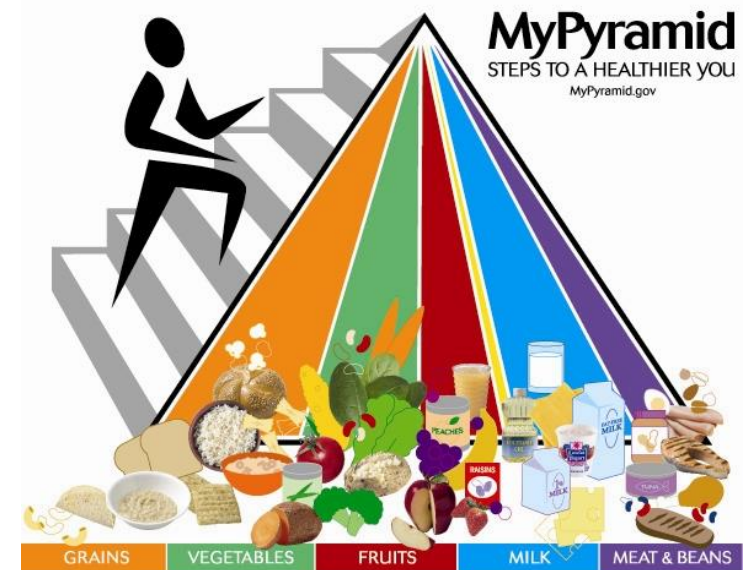
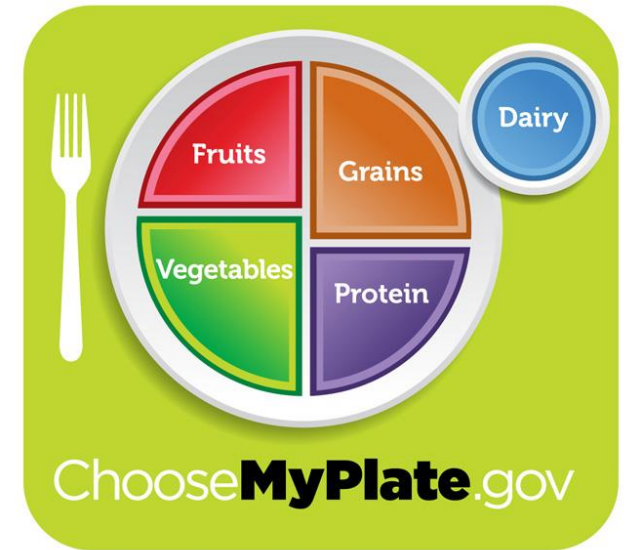
Reporting to students

Reporting to parents

# Health

One of the most iconic data visualizations in the world of health and nutrition is the food pyramid. In 2011, it changed to the food plate to make understanding your ideal diet easier.

In our "What Should I be Eating?" Activity, students compare the two visualizations, note pros and cons of each, and then produce ideas for an improved version.



# English/Literature

Probably one of the most unsuspecting places to find big data is in a book. When working as a team, students can catalogue words, themes, or phrases throughout the entire text. There are so many possibilities of how a novel can be visualized!

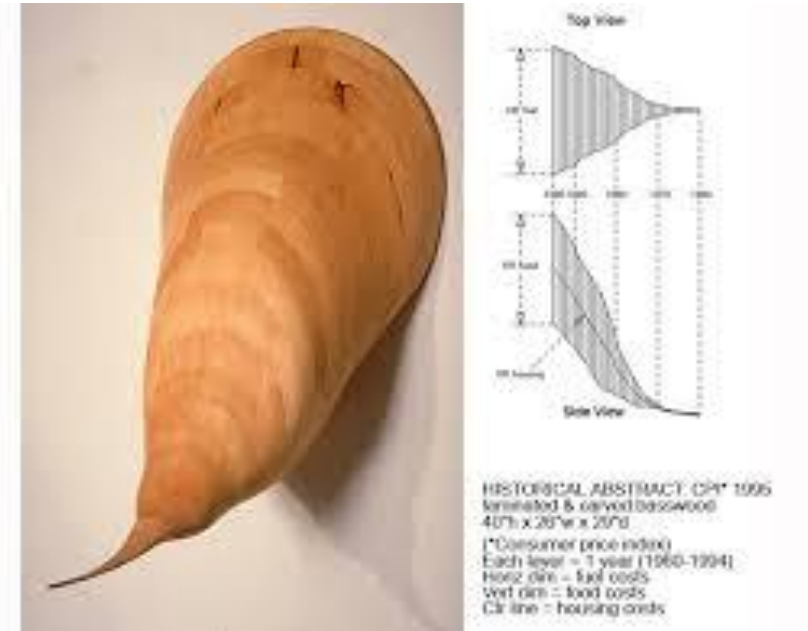
One teacher made a video of his students' work to analyze *The Iliad*. One of the activities in our book has students making a word chart on Infogram based on the frequency of words and themes in *The Great Gatsby*.



# Art

This subject gives the most freedom to young data scientists. While not the most practical in a business setting, data art can have a very strong emotional impact. These are some of Loren Madsen's pieces, the top is *CPI/Cost of Living* and the bottom is *6,000,000,000 monkeys* (human population from 10,000 BCE to 2000 CE).

There are several activities in our book that help students to define themselves through data, and one that shows them how to take raw data and turn it into wearable objects.

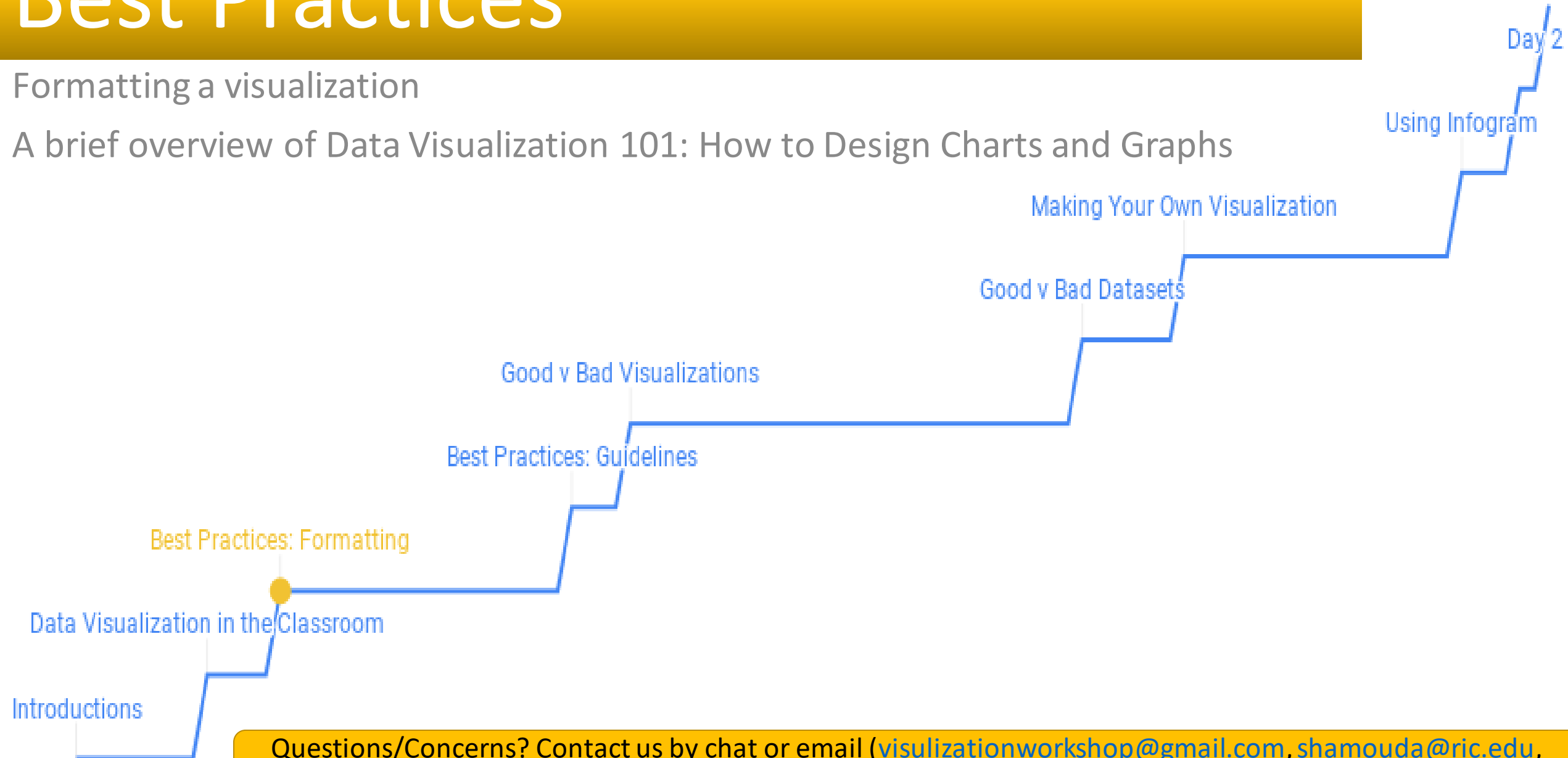


The sculpture represents world population from 10,000 BCE to 2000 CE. Half of all bones (up to over burial) are alive today.

# Best Practices

Formatting a visualization

A brief overview of Data Visualization 101: How to Design Charts and Graphs



Questions/Concerns? Contact us by chat or email ([visulizationworkshop@gmail.com](mailto:visulizationworkshop@gmail.com), [shamouda@ric.edu](mailto:shamouda@ric.edu), [lcendella\\_2763@email.ric.edu](mailto:lcendella_2763@email.ric.edu))

# Before you make a visualization

- Understand the data story you are *trying* to tell
  - What you want to see and what the data shows can be two very different things

I have data on participation in extracurricular activities and test performance.

I want to show that spending time on the former negatively impacts the latter.



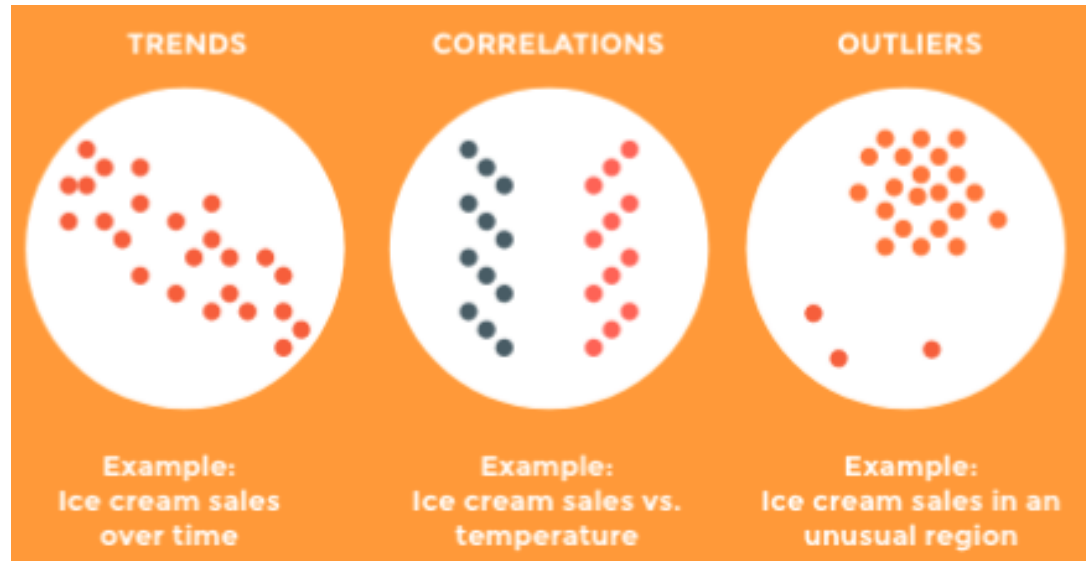
I want to see if spending time on the former correlates to the latter.

# Before you make a visualization

- Understand what type of data you are looking at the relationships between each other
  - Try to look for trends, correlations and outliers
  - Data can be quantitative, continuous, discrete, or categorical
  - Relationships can be nominal comparison, time series, correlation, ranking, deviation, distribution, part-to-whole
- What does all of this mean?



# Patterns and Data Types



## QUANTITATIVE

Data that can be counted or measured; all values are numerical.



## DISCRETE

Numerical data that has a finite number possible values. Example: Number of employees in the office.



## CONTINUOUS

Data that is measured and has a value within a range. Example: Rainfall in a year.



## CATEGORICAL

Data that can be sorted according to group or category. Example: Types of products sold.



# Bar Chart

Best used to show change over time, compare different categories, or compare parts of a whole

- Space between bars should be  $\frac{1}{2}$  the bar width
- Start the Y-axis at 0
- Use one color; you can have 1 accent color
- Order categories alphabetically, sequentially, or by value
- Use horizontal labels

## CONTENT PUBLISHED, BY CATEGORY

Educational Information



Entertainment



Thought Leadership



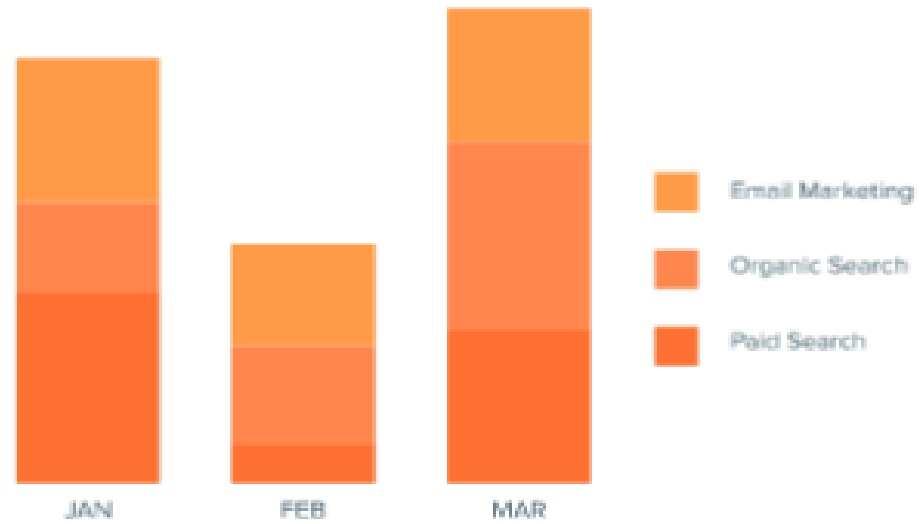
Horizontal charts best for  
long category names

Vertical charts best for  
chronological data or  
when values are  
negative

## PAGE VIEWS, BY MONTH



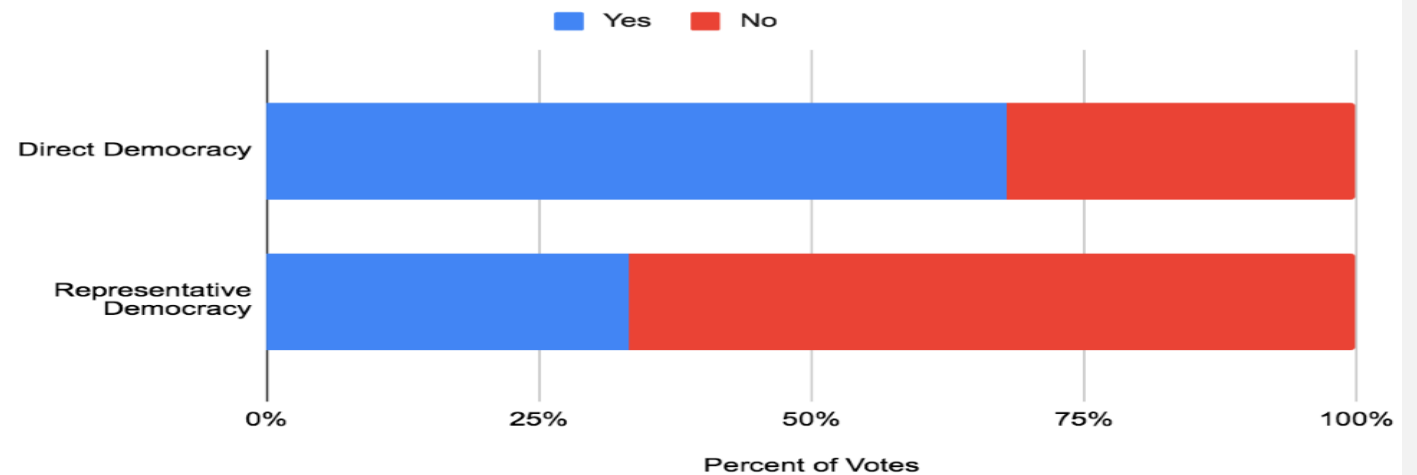
## MONTHLY TRAFFIC, BY SOURCE



Stacked charts compare part-to-whole relationships, and can use discrete or continuous data

100% Stacked charts show the percent distribution of categories. This figure comes from our "Voting: Then and Now" activity.

## Percent of Votes by Democracy



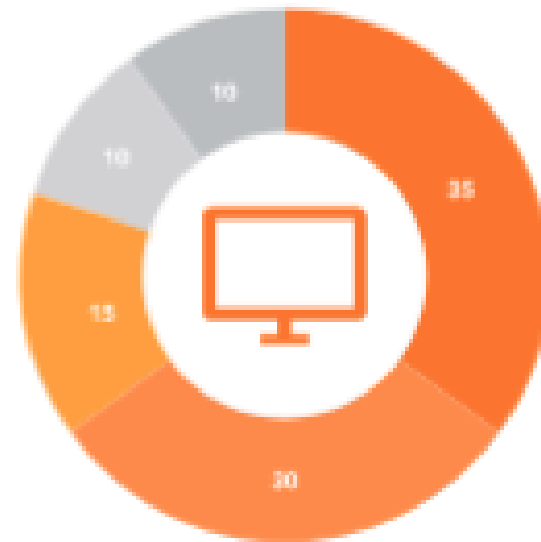
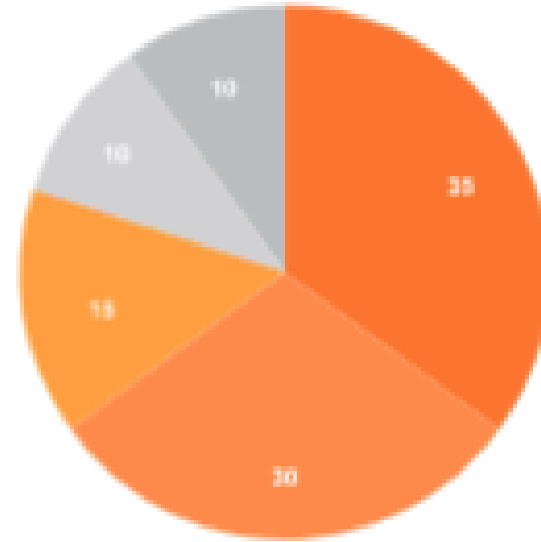
# Pie Chart

Best used for part-to-whole comparison with continuous or discrete data

- No more than 5 categories per chart
- Do not use pie charts for comparison
  - Stacked/100% stacked bar charts work better
- Make sure the data adds up to 100%
- The largest slice goes at 12 o'clock; all remaining slices go clockwise or counterclockwise in order of decreasing value
  - i.e. the 2nd largest value would go next, then 3rd largest, etc.

## Pie Charts

A standard pie chart and a donut chart. They show the same data, but the latter can have a stylistic element in the middle.

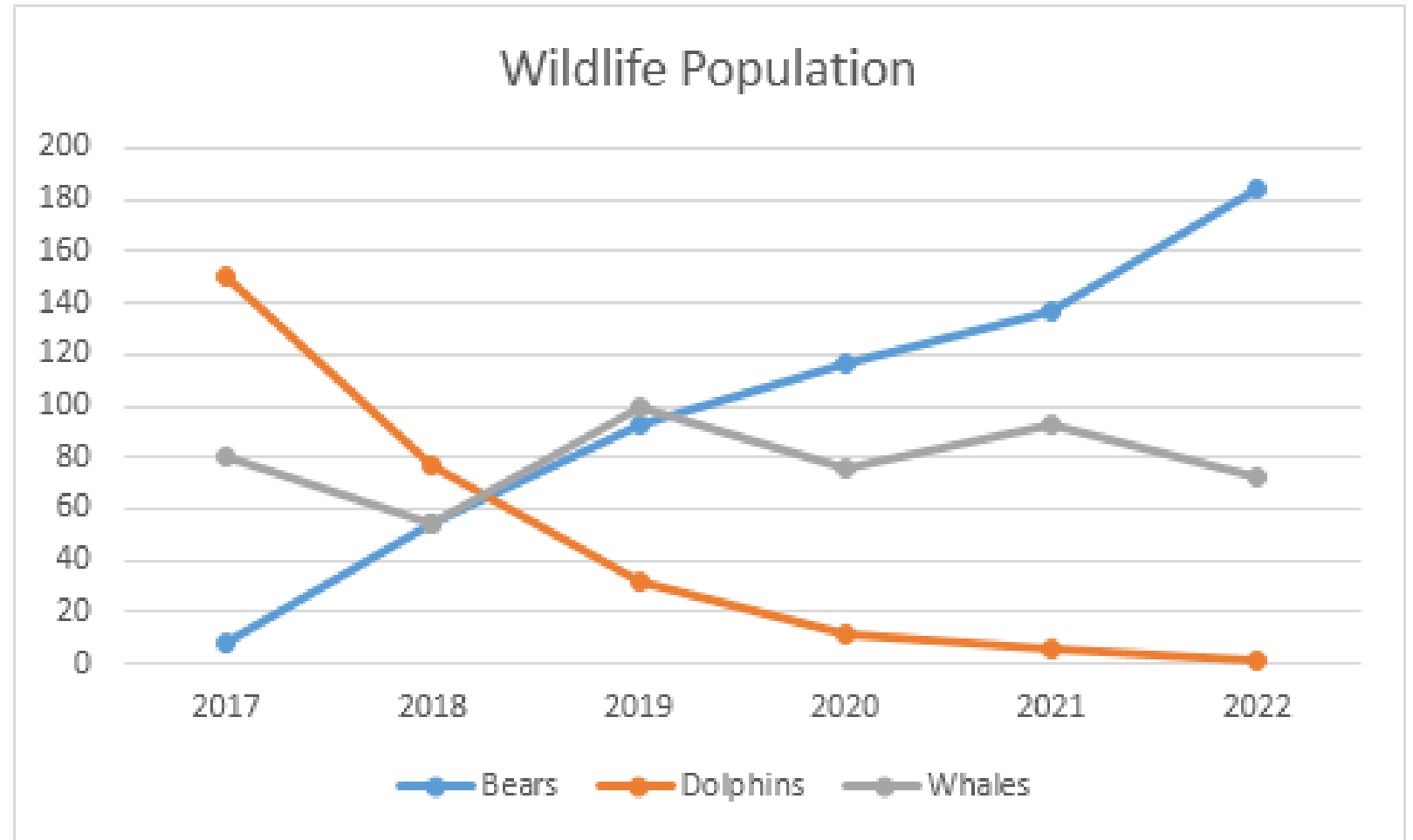


# Line Chart

Best used to show time-series relationships with continuous data

- Start Y-axis at 0, if possible
- No more than 4 lines
- Use solid lines
- Label lines directly (instead of separate legend)
- Line chart should take up two-thirds of Y-axis' scale

# Line Chart



# Scatter Plot

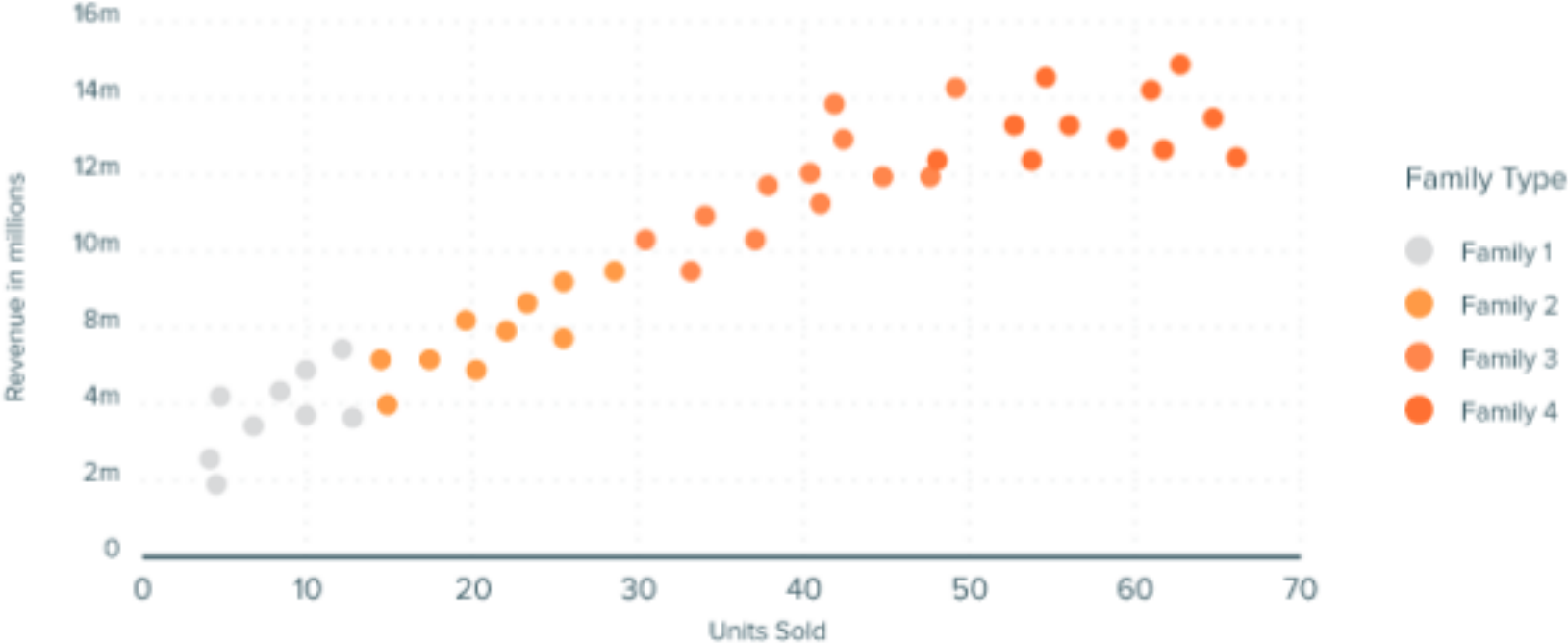
Best used to show correlation in large amounts of data

- Start Y-axis at 0
- Use size and dot color to encode more variables
- Use trendlines, but no more than 2



# Scatter Plot

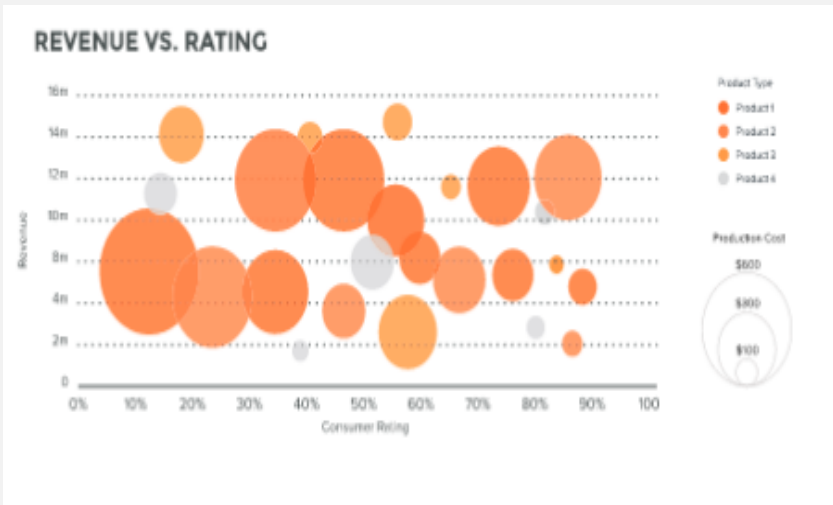
## REVENUE, BY PRODUCT FAMILY



# Bubble Chart

Best used for nominal comparisons or ranking relationships

- Make sure all bubble labels are visible
- Size bubbles according to area, not diameter!
- Don't use odd shapes



Bubble plots are really scatter plots with another variable being displayed

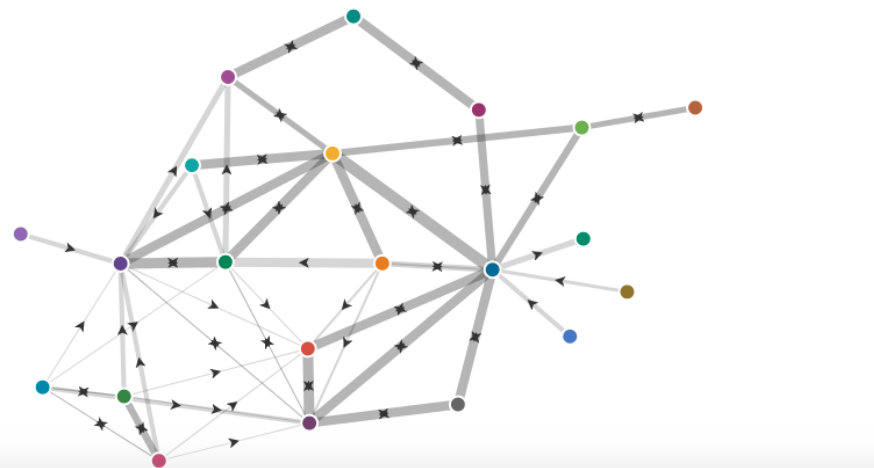


Bubble maps show values at specific regions

Node diagrams, an off-shoot of bubble charts, can show values as bubbles of different sizes, and how they are related to each other. This figure is a sample from our "Social Networking – Who Knows Who" English/Literature activity

**Romeo and Juliet Characters**

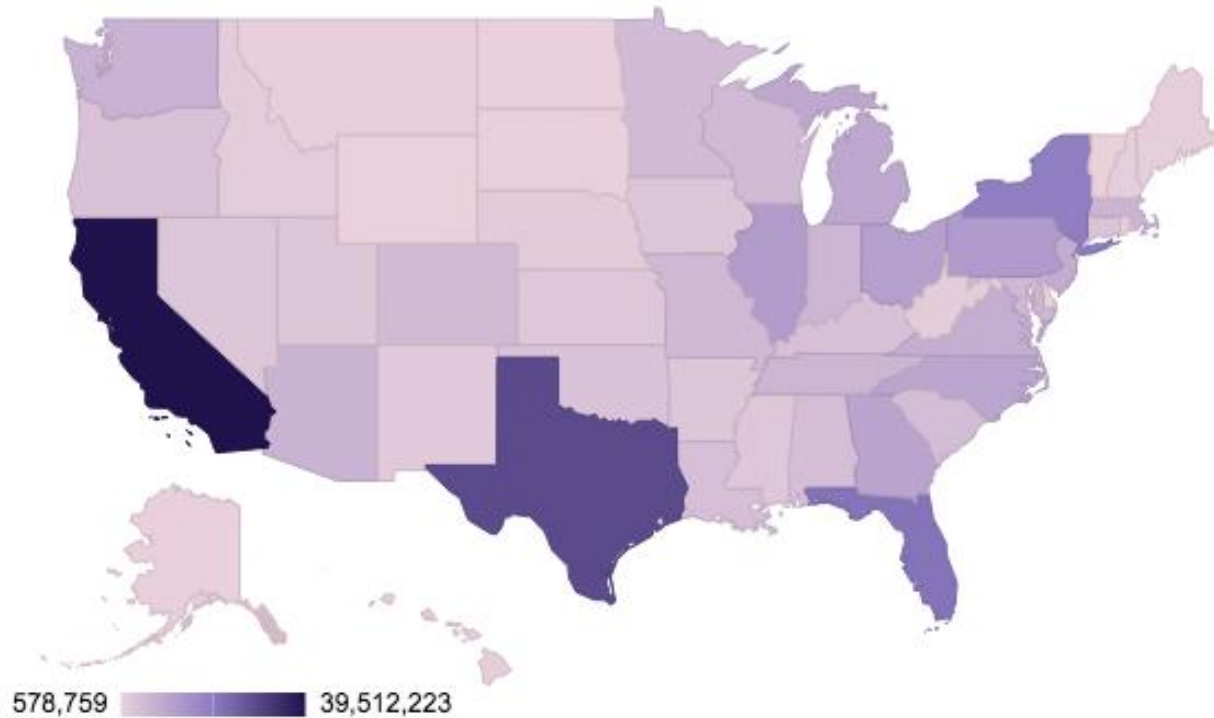
- Romeo
- Juliet
- Friar Lawrence
- Mercutio
- The Nurse
- Tybalt
- Capulet
- Lady Capulet
- Montague
- Lady Montague
- Paris
- Benvolio
- Prince Escalus
- Friar John
- Balthasar
- Sampson
- Gregory
- Abram
- The Apothecary
- Peter
- Rosaline
- The Chorus



# Heat Map

Best used for categorical data and values with color intensity

- Use a simple map outline
- Use shades of a single color, or a spectrum of two analogous colors to show intensity
- Intuitively code color intensity to values
  - i.e. don't make light colors represent large values
- Don't use more than 1 pattern overlay
- Choose appropriate data ranges. Use +/- to extend data ranges

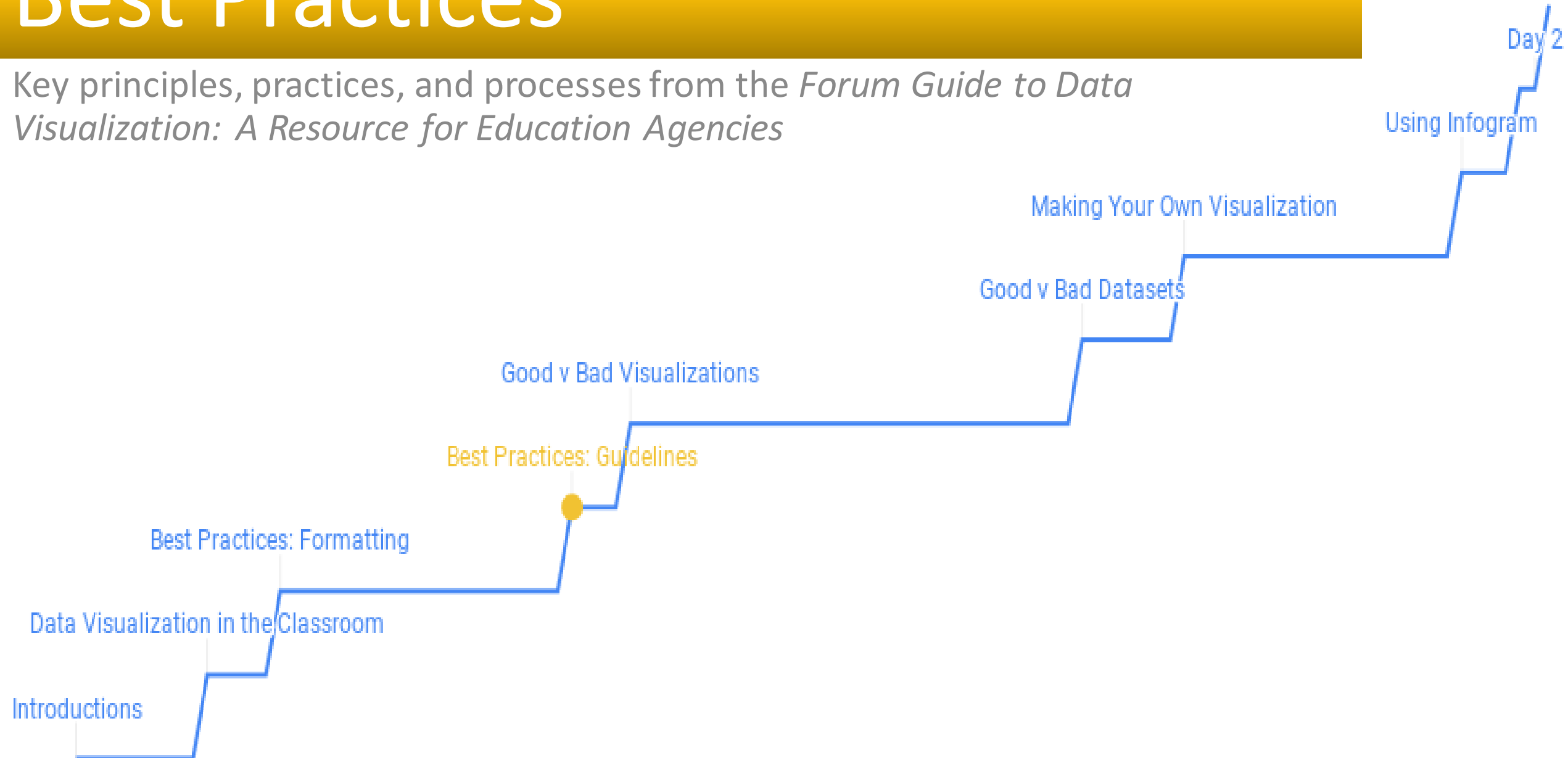


The example heat map made from the "Mapping America" activity in our activity book. Students navigate the U.S. Census Bureau to find population data, find the data story they want to tell, and then create a coherent map in Google Sheets.

This map uses a color range of light pink to dark purple. I felt the light pink was better than light purple to help distinguish between low and moderate levels of population.

# Best Practices

Key principles, practices, and processes from the *Forum Guide to Data Visualization: A Resource for Education Agencies*



# Principles

## 1: Show the data

- Including correct labels and metadata

## 2: Reduce the clutter

## 3: Integrate text and images

- Visualization must be able to stand alone and reinforce text

## 4: Portray data meaning accurately and ethically

- No cherry-picking or suggesting conclusions where there aren't any

# Practices

1: Capitalize on consistency

2: Data that shouldn't be compared shouldn't be presented side by side

3: Don't limit design choices to default graphing programs

4: Focus on the take-home message for the audience

5: Minimize acronyms, jargon, and technical terms

6: Choose a font that's easy to read and reproduces well

7: Recognize the importance of color



# Processes

1: Question – Someone needs information

2: Research – Data exploration and analysis

3: Findings – Data meaning/answer

4: Customization – Audience-specific messaging

5: Visualization – Present data meaning clearly and accurately

6: User Feedback – Review and refine efforts

# 10-minute break

## Pre-workshop notes



Use headphones



Everyone their own laptop



Quiet space



Reliable internet connection



Camera on

## online etiquette



Mute yourself



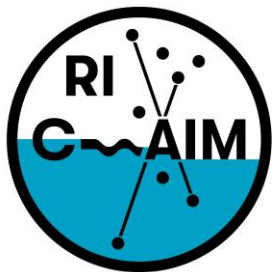
Press spacebar to unmute



Want to speak



Yes / agree



RHODE ISLAND CONSORTIUM FOR  
Coastal Ecology  
Assessment  
Innovation &  
Modeling

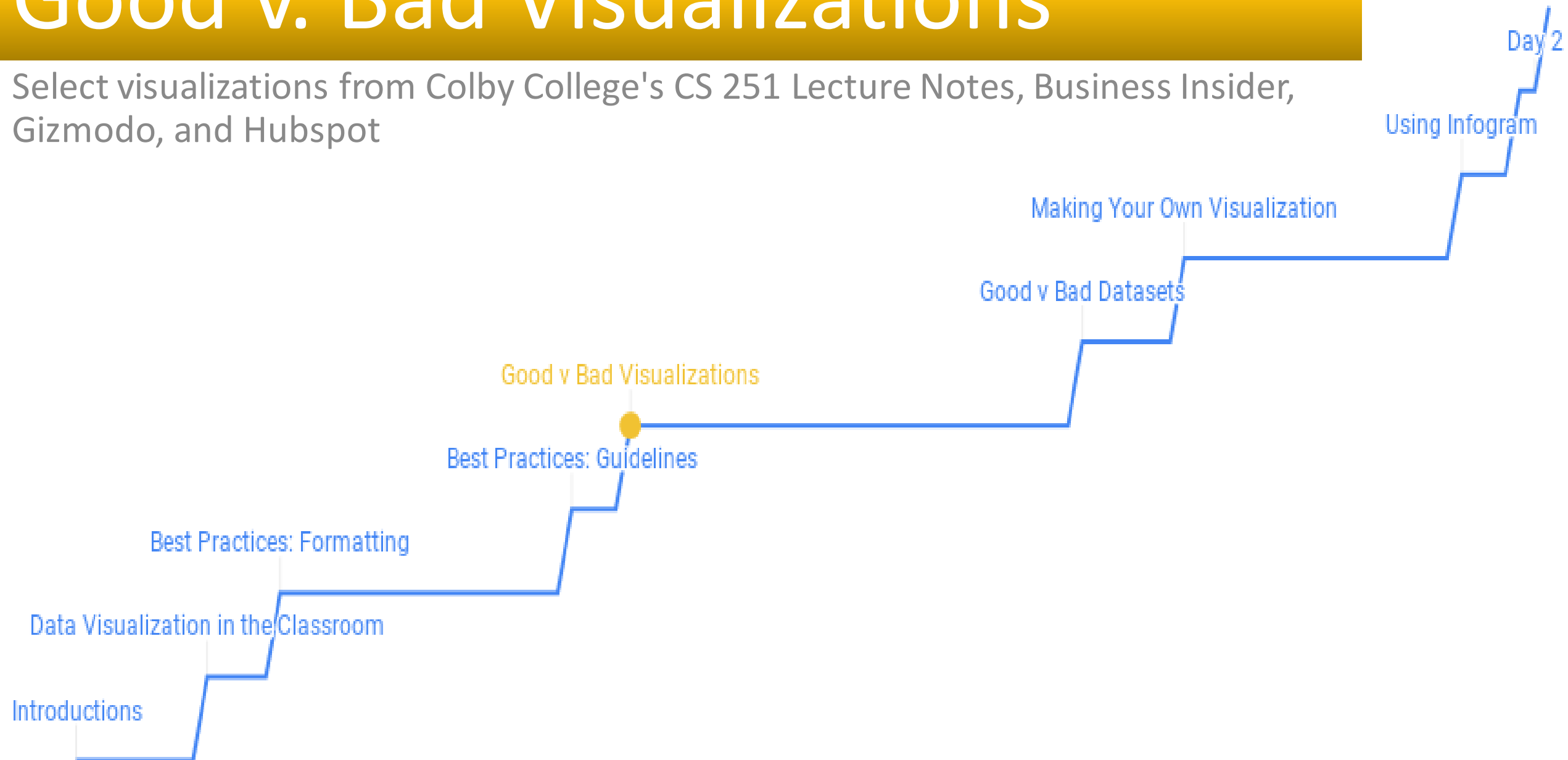


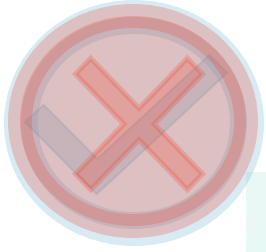
This material is based upon work supported in part by the National Science Foundation under EPSCoR Cooperative Agreement #OIA-1655221

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

# Good v. Bad Visualizations

Select visualizations from Colby College's CS 251 Lecture Notes, Business Insider, Gizmodo, and Hubspot



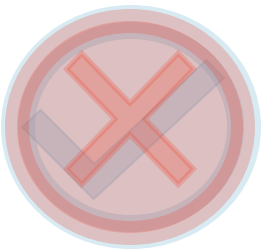


- Using icons as bars implies that area is something to consider
  - Burger King sold 3x more than Starbucks but has an area 9x greater
- While it should have a title and metadata, the original infographic it came from does include this information

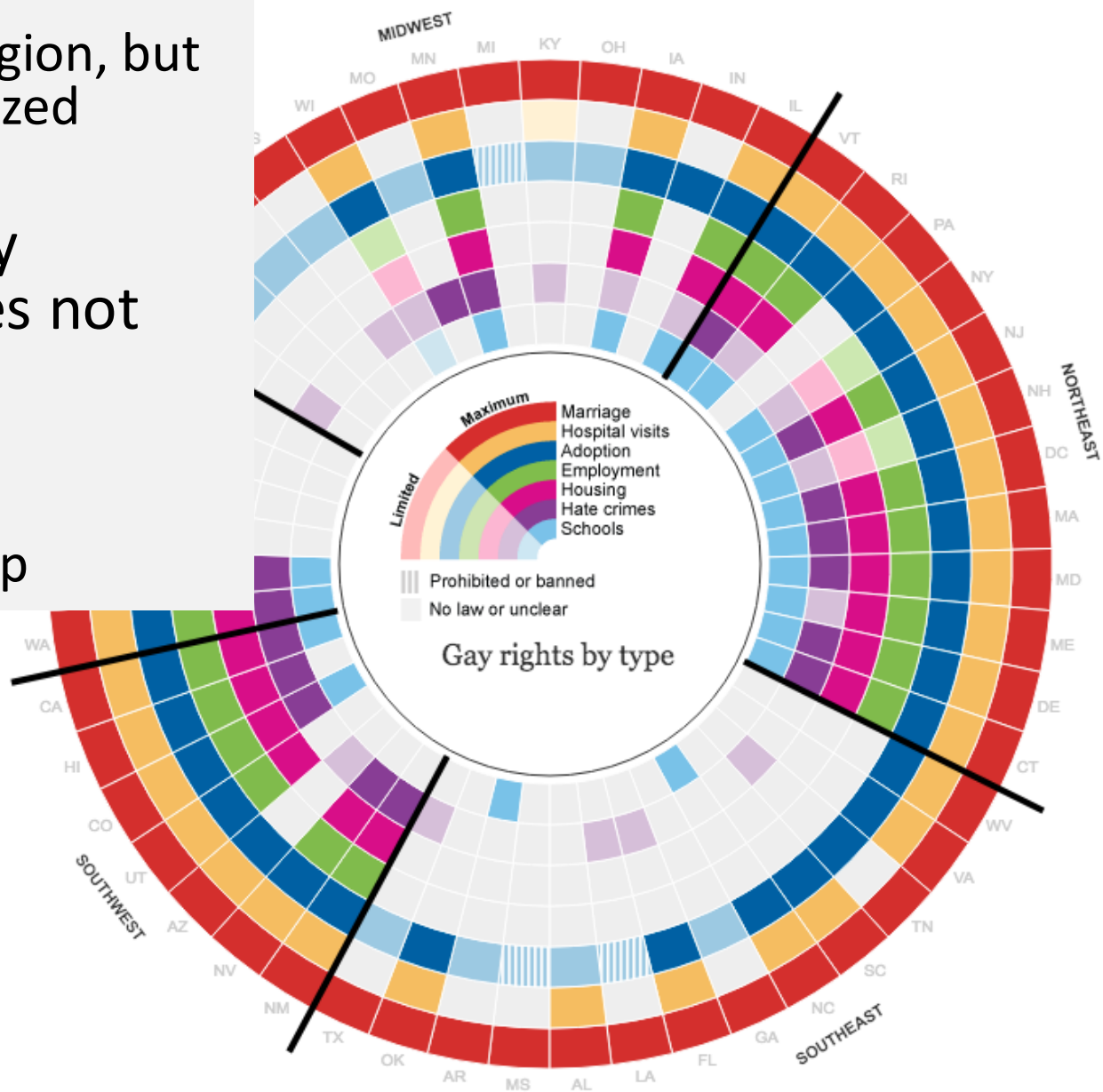


Fast Food Chain Sales – bar chart

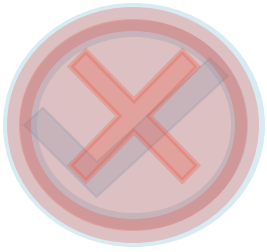




- Not very well organized
  - States are grouped by region, but not alphabetically organized within
- Has an option to scale by population, but that does not seem to be relevant
  - Laws do not correlate to population
  - Some state don't show up



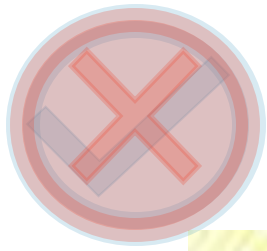
Gay Rights by Type  
(2015) - donut chart



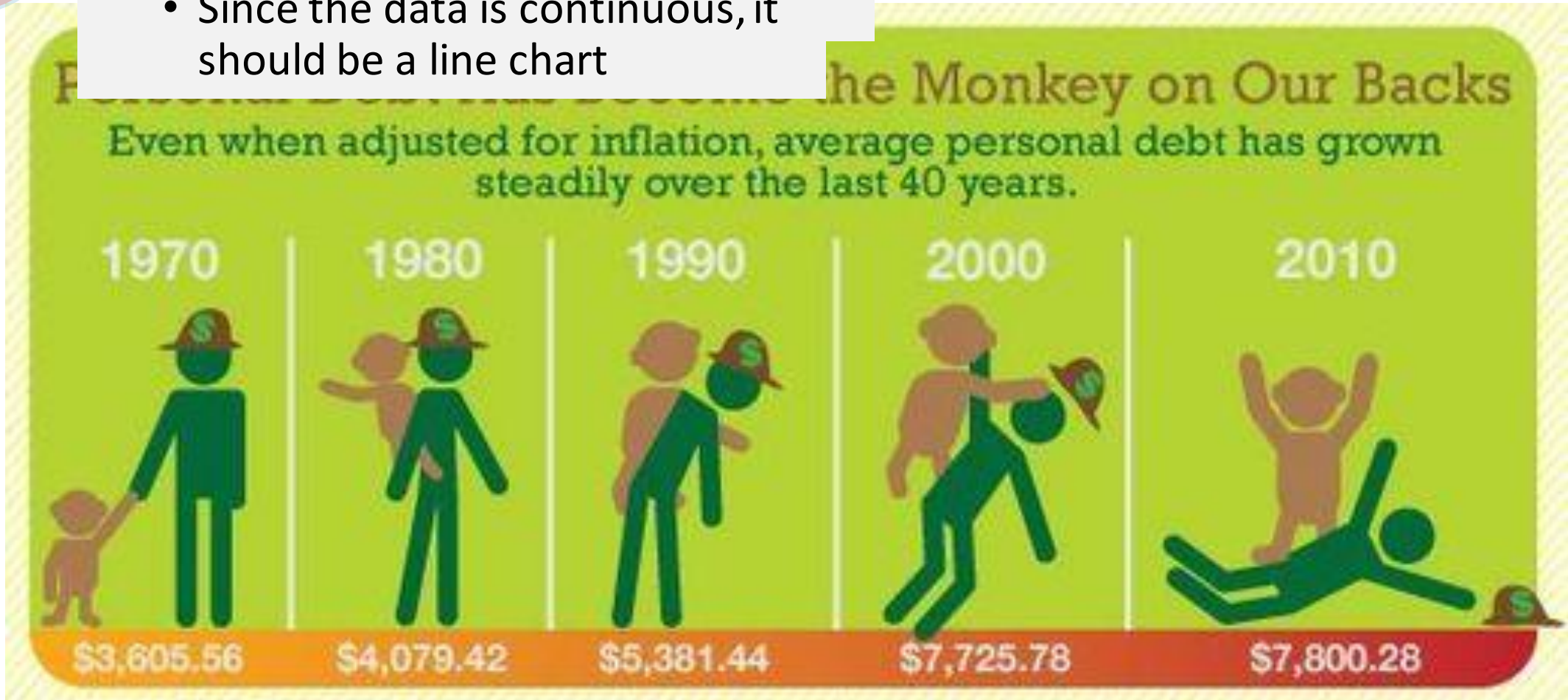
- No title, legend, scale, metadata
  - Going to their website, the graph has been replaced by a table
  - All of the states were within 2 percentage points of the national average (3.5%) so every state became green

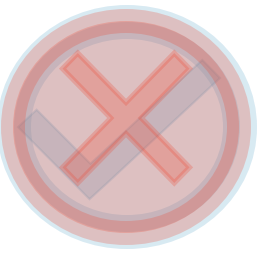
## LGBT Percentage by State – heat map



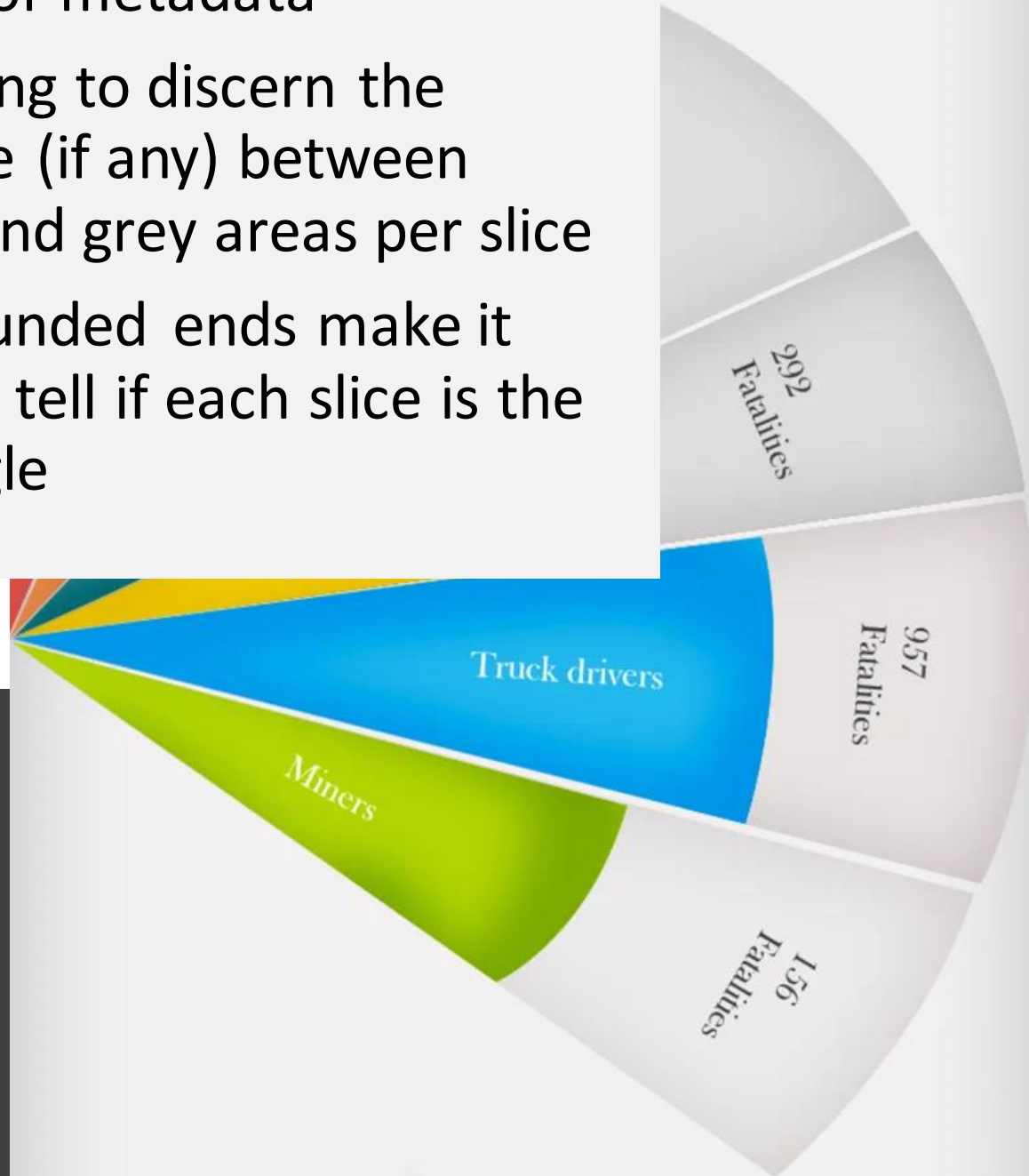


- Monkeys are not a valid chart element
  - Since the data is continuous, it should be a line chart





- Not title or metadata
- No marking to discern the difference (if any) between colored and grey areas per slice
- Oddly-rounded ends make it harder to tell if each slice is the same angle

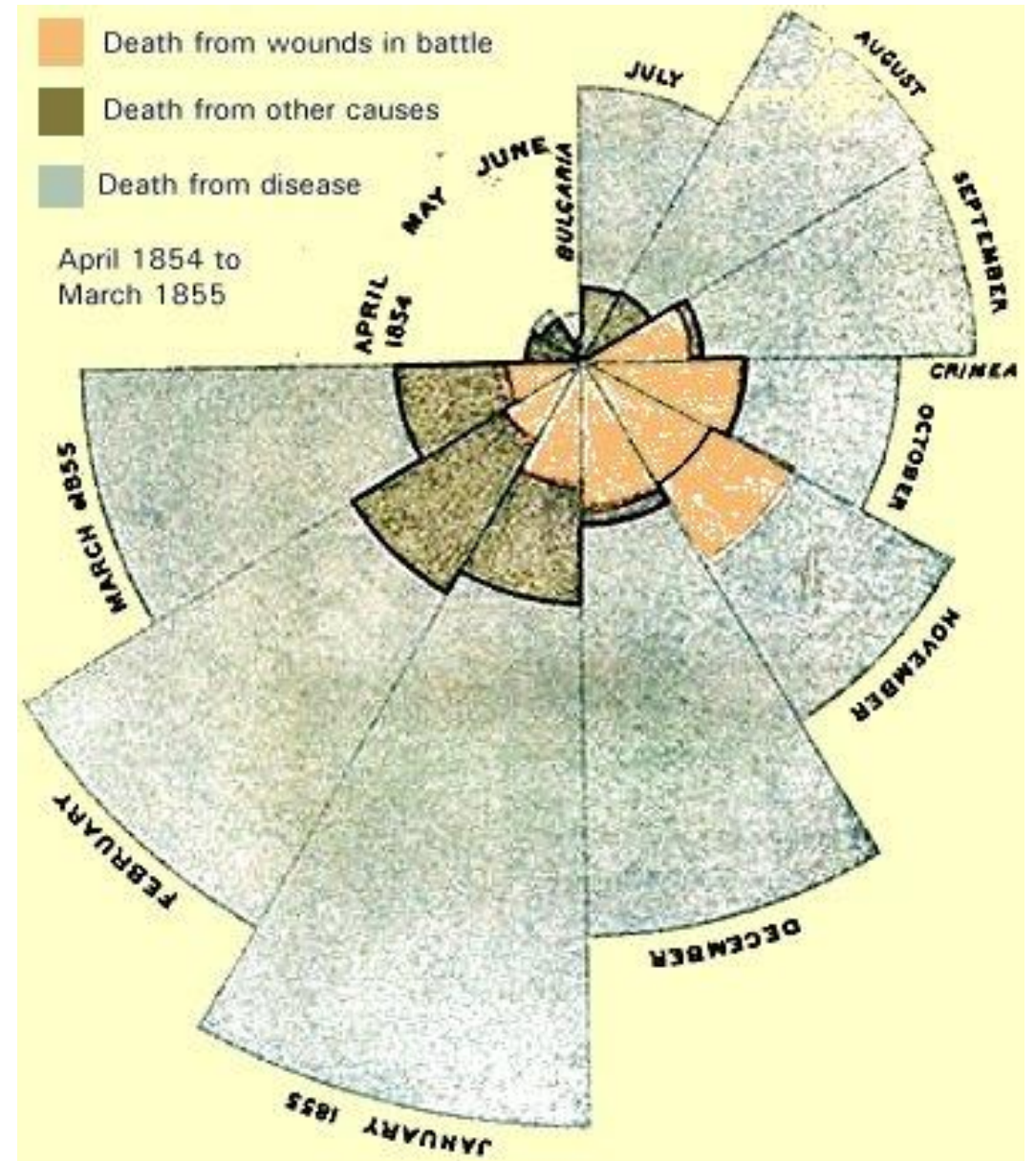


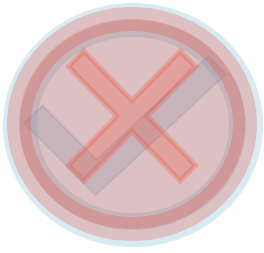
Fatalities by Occupation – coxcomb chart?



# A note on coxcombs

- Attributed to Florence Nightingale, coxcombs are an off-shoot of pie charts
  - Instead of slice size representing data, the slice length is what's important
- They are not good for data sets with small differences
  - Better to show clear patterns, which is what she did (how infection is what really kills soldiers in war)





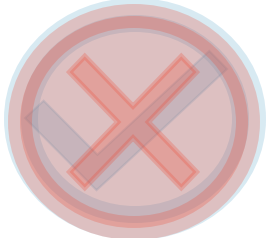
- There's no color and it wasn't greyscaled!
  - Remember to check out what something will look like when you print. There should be a monochrome option in the printer preview
  - Remember to think about colorblind students, as well. Even with color, they may not be able to see differences



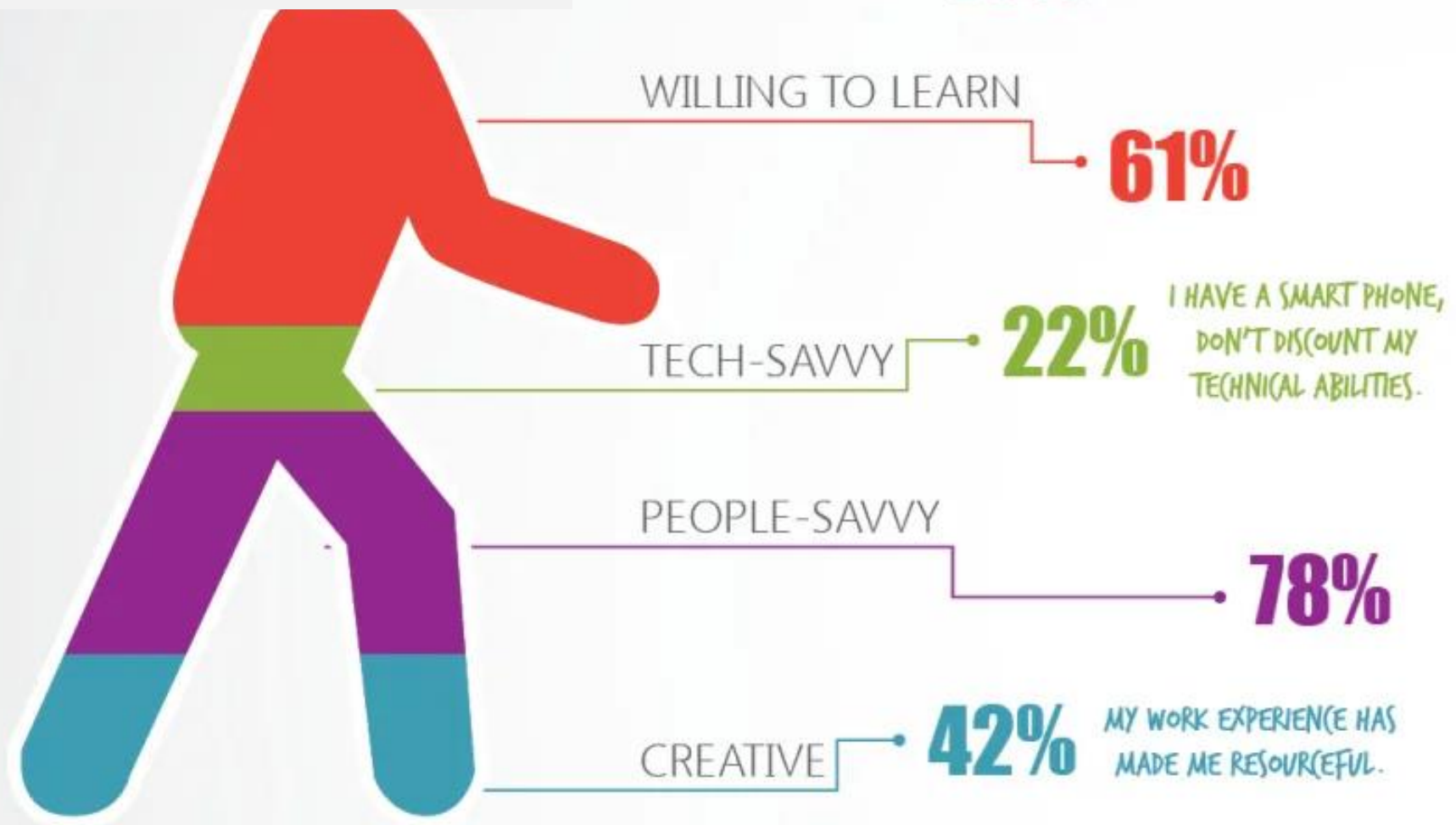
heat map



# HOW BABY BOOMERS DESCRIBE THEMSELVES

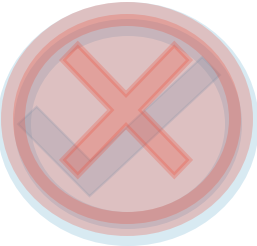


- Odd shape leads to confusion
  - Are we looking at area or height?
- Percentages don't add up to 100



100% stacked bar chart





- Hard to tell how long they spent on activities
  - No start/end time for each segment

over a segment

- Has useful sorting to compare each person's schedule by activity

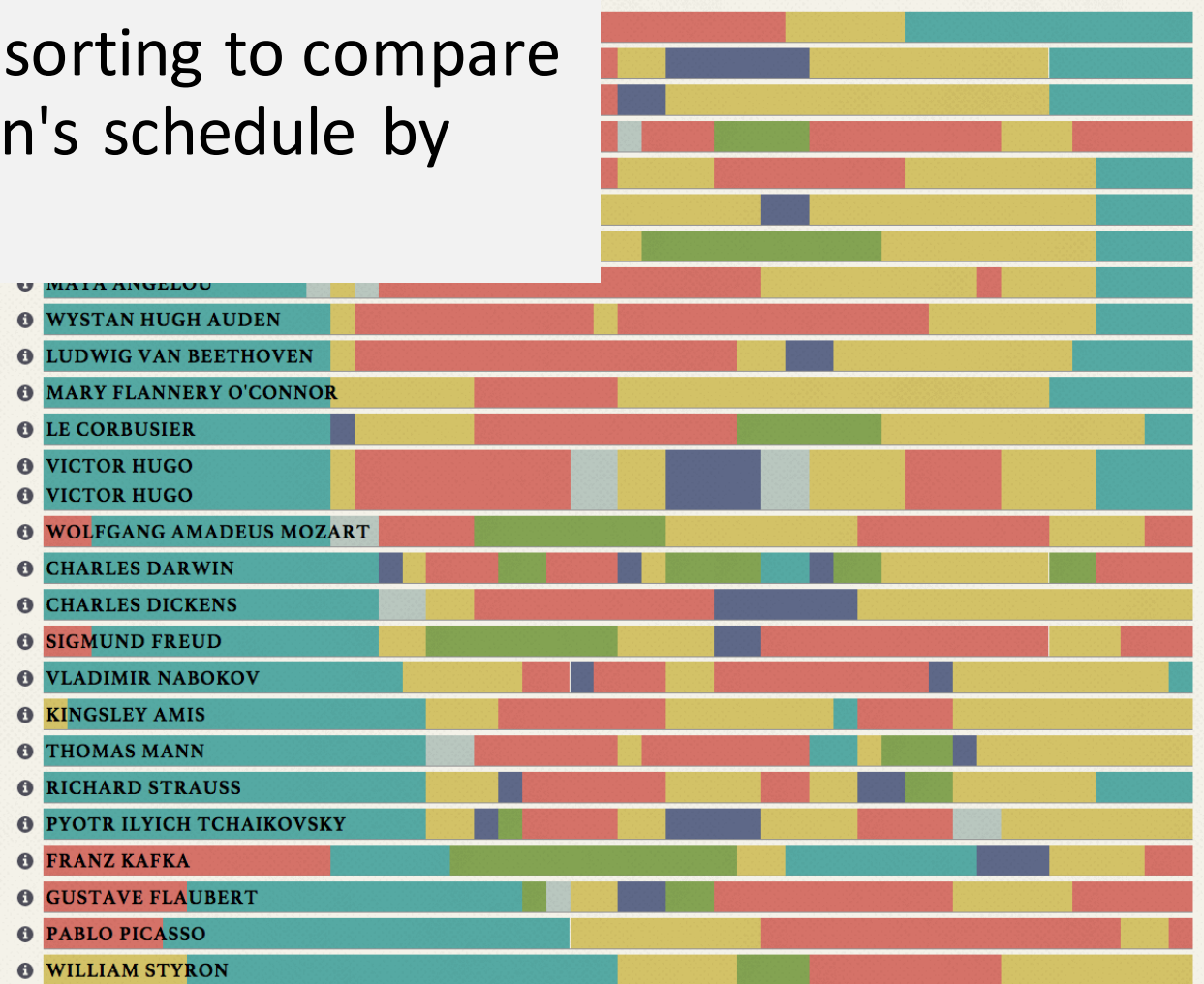
# STATISTICS OF

PEOPLE

Discover how some of the world's most famous people structured their day, based on different categories by toggling on and off. Click on a name to learn more about the daily schedule of that person.

LEISURE EXERCISE OTHER

12 PM 1 2 3 4 5 6 7 8 9 10 11 12



100% stacked bar chart



What common chart type does it try to reflect?

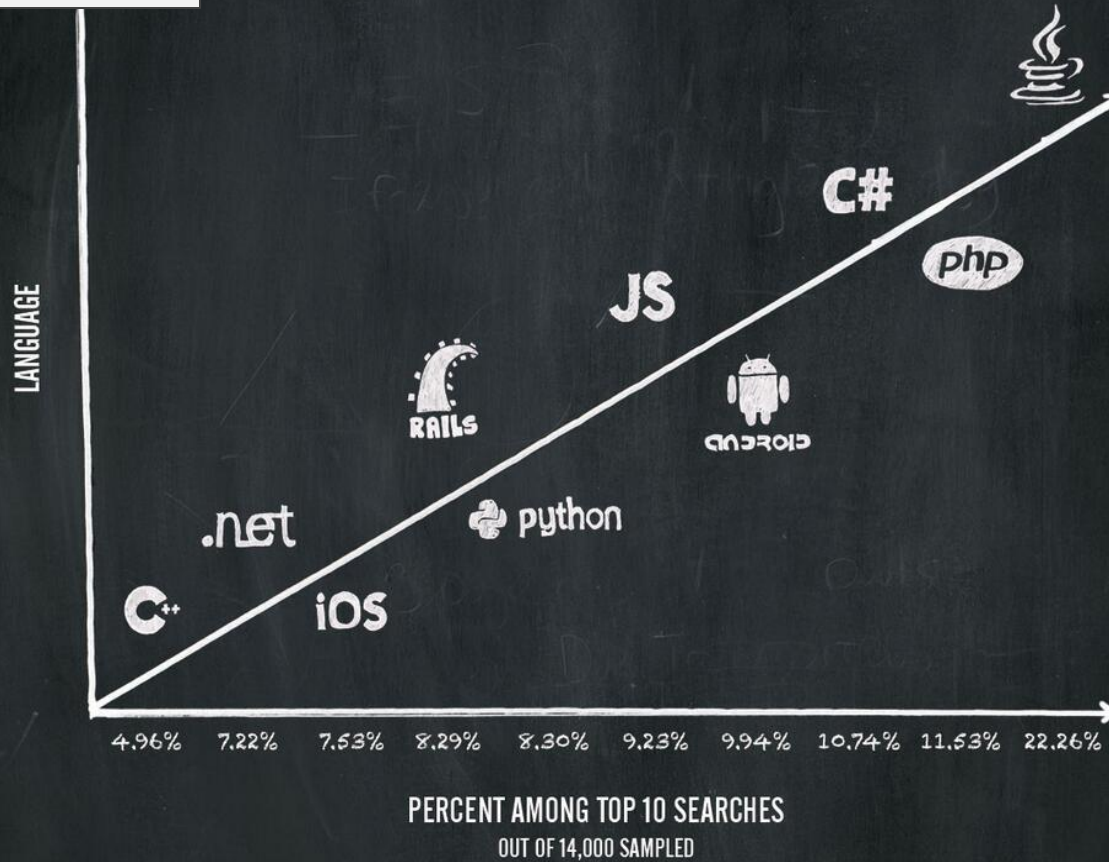
What do you think is wrong with this graph?

What do you think is right with this graph?

readwrite presents

# 10 MOST IN DEMAND DEVELOPER SKILLS OF 2013

INFORMATION COMPILED BY STACK OVERFLOW



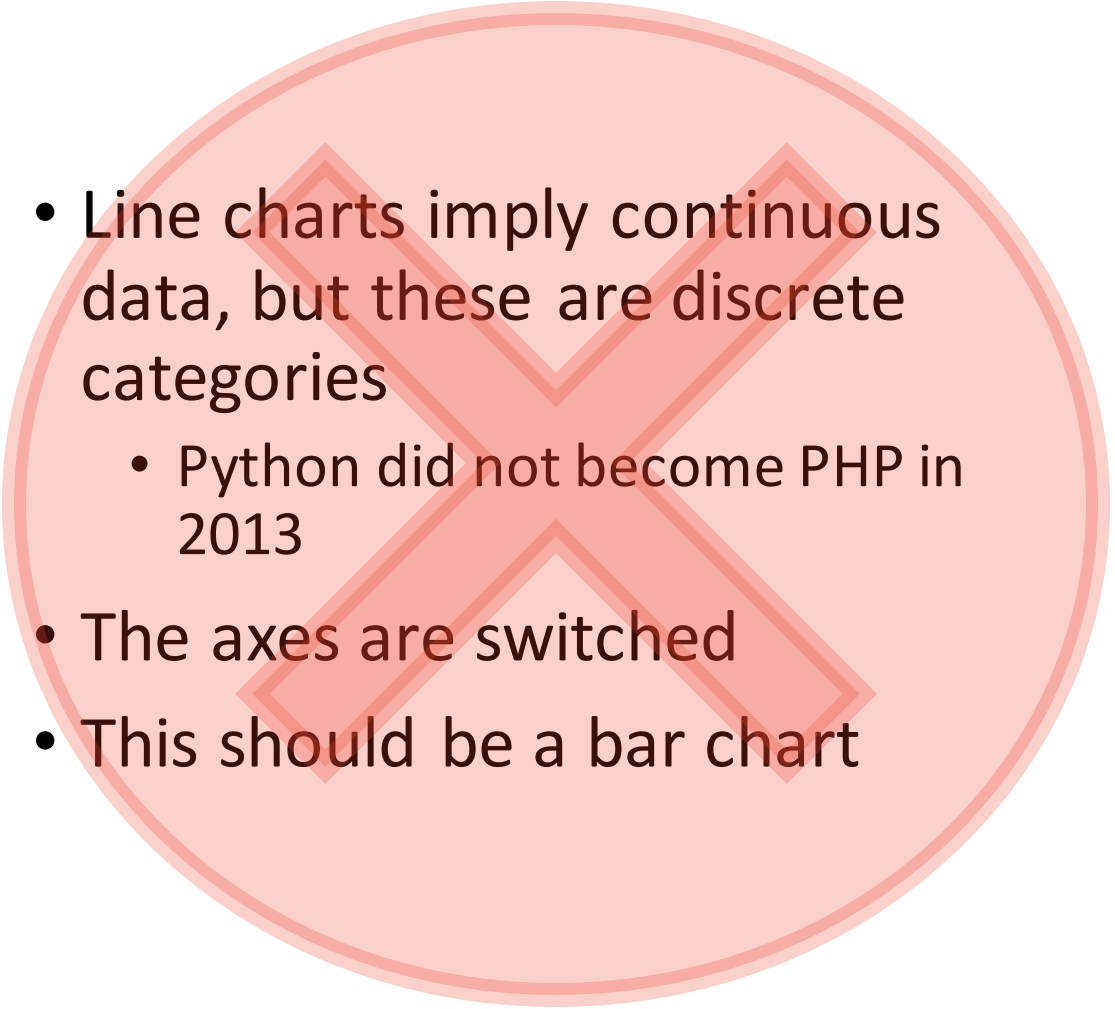
Go to [menti.com](https://www.menti.com) and type in the code or follow  
this link

<https://www.menti.com/vpi64mrssw>

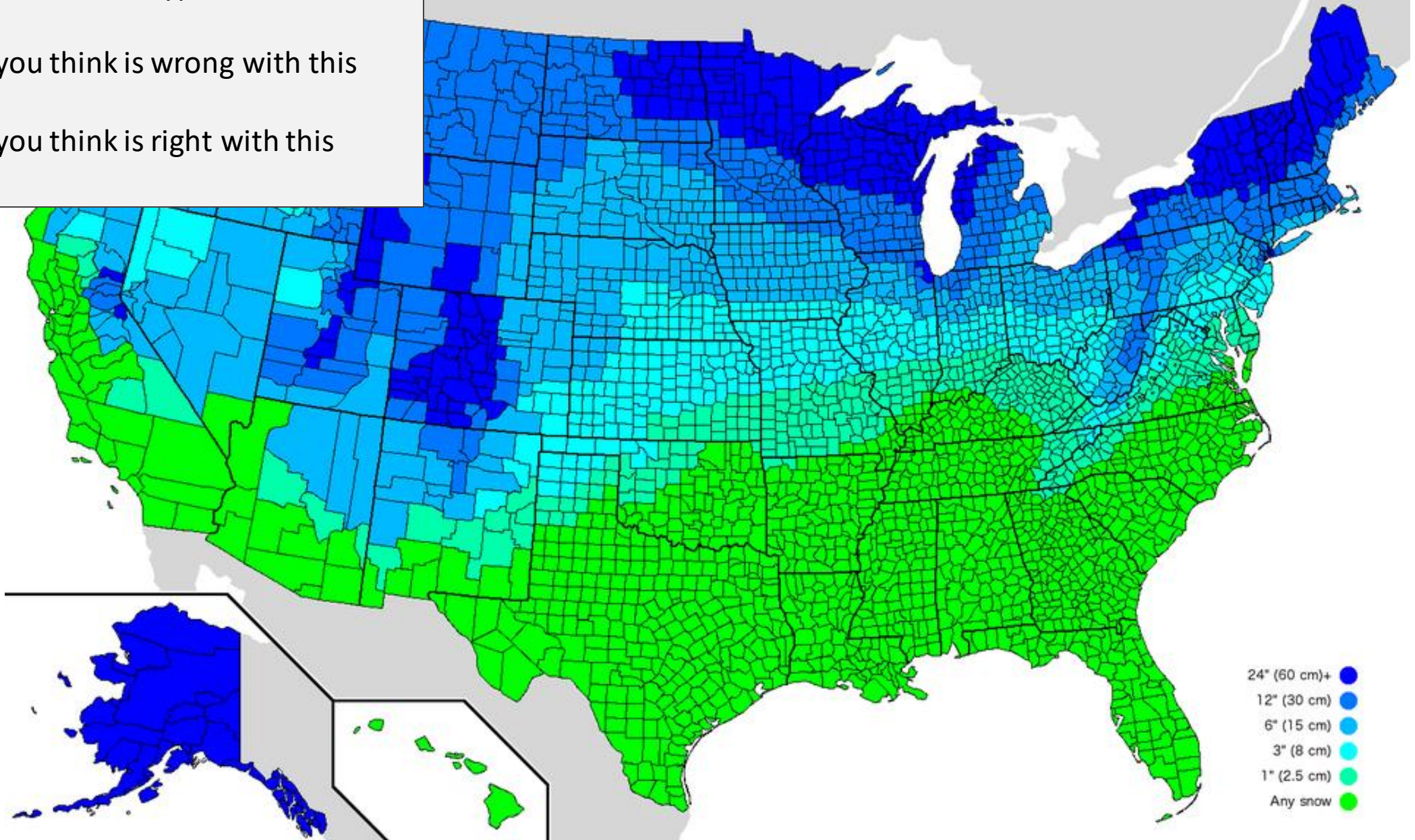
**\*\*\*\*\*Mentimeter questions\*\*\*\*\* Link: 14 58 7**

# In Demand Developer Skills – line chart

- 
- Has title, metadata

- 
- Line charts imply continuous data, but these are discrete categories
    - Python did not become PHP in 2013
  - The axes are switched
  - This should be a bar chart

What common chart type does it reflect?  
What do you think is wrong with this graph?  
What do you think is right with this graph?



How Much Snow It Typically Takes to Cancel School in the U.S.






Go to [menti.com](https://www.menti.com) and type in the code or follow  
this link

<https://www.menti.com/k541p6tn8a>

**\*\*\*\*\*Mentimeter questions\*\*\*\*\* Link: 69 57 58**

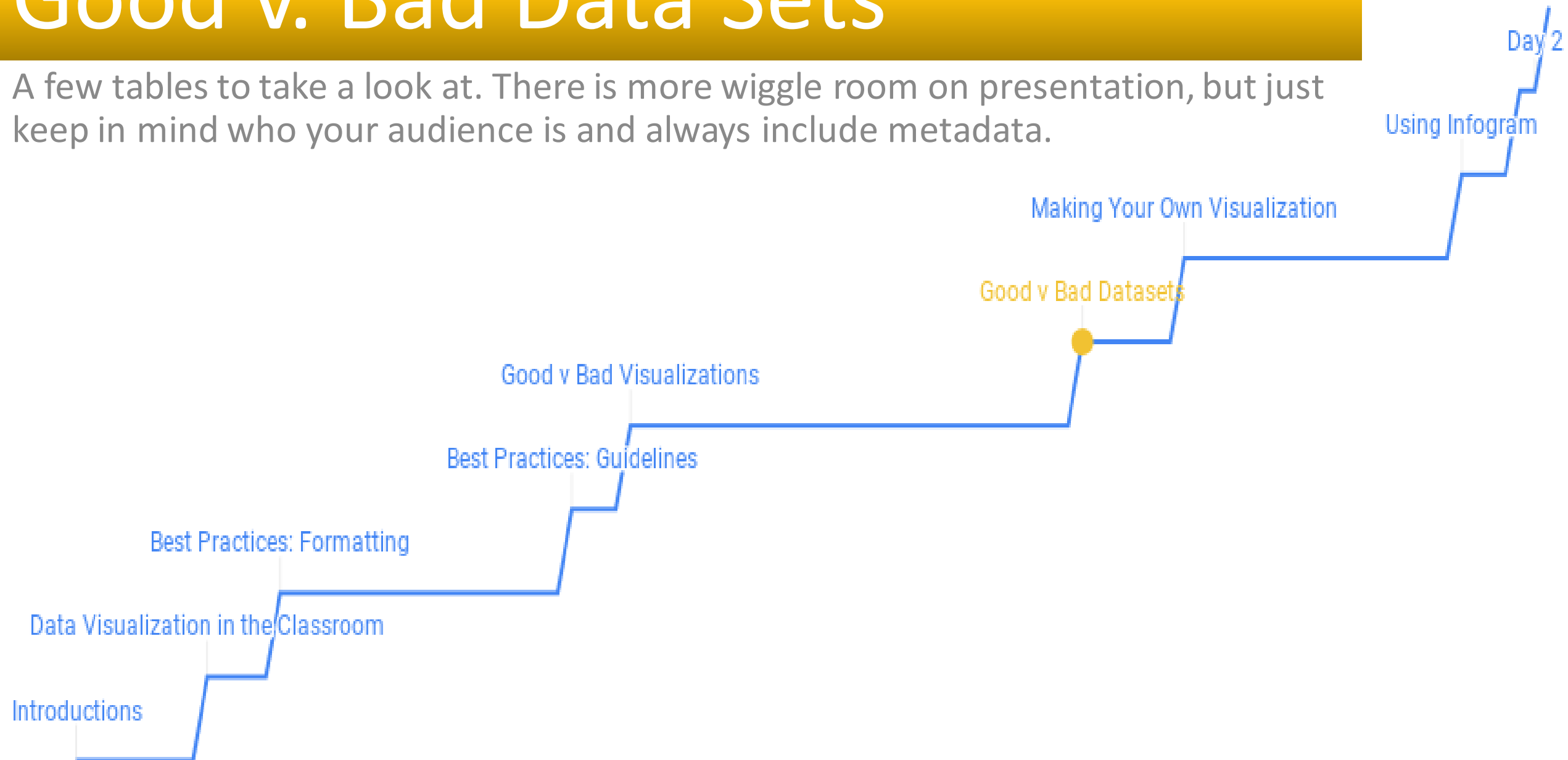
# How Much Snow It Typically Takes to Cancel School in the U.S. - heat map

- 
- Beautiful color scheme and use of a district map
  - Has a legend
  - Has metadata (if you click the link)

- 
- No title

# Good v. Bad Data Sets

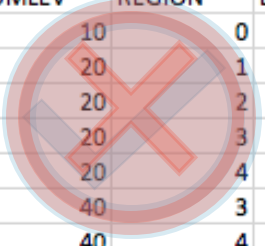
A few tables to take a look at. There is more wiggle room on presentation, but just keep in mind who your audience is and always include metadata.





# U.S. Census Bureau Annual Population Estimates

SUMLEV	REGION	DIVISION	STATE	NAME	CENSUS2010	ESTIMATESB	POPESTIMAT	POPESTIMAT	POPESTIMAT	POPESTIMAT	POPESTIMAT	POPESTIMAT	POPESTIMAT	POPESTIMAT	POPESTIMAT	POPESTIMAT	NPOPCHG	
10		0																
20		1																
20		2																
20		3																
20		4																
40		3																
40		4																
40		4																
40		4																
40		3																
40		4																
40		4																
40		1																
40		3																
40		3																
40		3																
40		3																
40		4	9	15	Hawaii	1360301	1360307	1363963										
40		4	8	16	Idaho	1567582	1567657	1570746										
40		2	3	17	Illinois	12830632	12831572	12840503										
40		2	3	18	Indiana	6483802	6484051	6490432										
40		2	4	19	Iowa	3046355	3046871	3050745										
40		2	4	20	Kansas	2853118	2853123	2858190										
40		3	6	21	Kentucky	4339367	4339333	4348181										
40		3	7	22	Louisiana	4533372	4533487	4544532										
40		1	1	23	Maine	1328361	1328358	1327629										
40		3	5	24	Maryland	5773552	5773794	5788645										
40		1	1	25	Massachusetts	6547629	6547785	6566307										



- There are some extra rows in there that aren't really mentioned until you try to visualize the data
  - Regions (north, south, midwest, west) are present in this data about state populations

0	315993715	318301008	320635163	322941311	324985539	326687501	328239523	563561
6	55901806	56006011	56034684	56042330	56059240	56046620	55982803	61691
13	67560379	67745167	67860583	67987540	68126781	68236628	68329004	44691
18	118364400	119624037	120997341	122351760	123542189	124569433	125580448	303650
13	74167130	74925793	75742555	76559681	77257329	77834820	78347268	153529
18	4830081	4841799	4852347	4863525	4874486	4887681	4903185	5312
13	737068	736283	737498	741456	739700	735139	731545	3661
18	6632764	6730413	6829676	6941072	7044008	7158024	7278717	14884
14	2959400	2967392	2978048	2989918	3001345	3009733	3017804	5933
10	38260787	38596972	38918045	39167117	39358497	39461588	39512223	64983
17	5269035	5350101	5450623	5539215	5611885	5691287	5758736	18030
17	3594841	3594524	3587122	3578141	3573297	3571520	3565287	4967

popchg2010-2019: Annual Estimates of Resident Population Change for the States, and Puerto Rico and State Rankings: April 1, 2010 to July 1, 2019

File: 7/1/2019 National and State Population Estimates

Source: U.S. Census Bureau, Population Division  
 Release Date: December 2019

Sort order of observations: States in FIPS code sort

Data fields (in order of appearance):

VARIABLE	DESCRIPTION
SUMLEV	Geographic summary level
REGION	Census Region code
DIVISION	Census Division code
STATE	State FIPS code
NAME	State name
ESTIMATESBASE2010	4/1/2010 resident total population estimates base

# U.S. Census Bureau Debt for Households

**Table 1. Median Value of Debt for Households, by Type of Debt and Selected Characteristics: 2004**

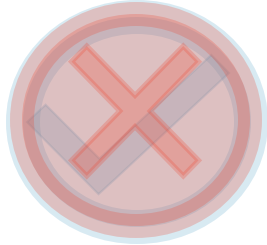
- Variable definitions aren't specified
  - "Loans" and "Other Debt" could mean anything

	Total Debt				Unsecured Debt				
	Business Debt	Vehicle Debt	Total	Credit Card Debt	Loans	Other Debt			
All households	10,000	11,000	50,000	3,000	5,500	7,000			
White Alone (Not of Hispanic Origin)	65,000	80,000	95,000	20,000	11,000	5,000	3,000	5,500	7,000
Black Alone	30,000	48,000	80,000	10,000	10,000	5,000	3,000	4,000	7,000
Asian Alone	90,000	120,000	150,000	(B)	12,000	5,000	3,000	(B)	11,000
Other (residual)	38,000	50,200	85,000	(B)	10,000	5,000	3,000	4,500	5,000
Hispanic Origin									
Not of Hispanic Origin									
<b>AGE OF HOUSEHOLD</b>									
Less than 35 years									
35 to 44 years									
45 to 54 years									
55 to 64 years									
65 years and over									
65 to 69 years									
70 to 74 years									
75 and over									

NOTE: In dollars. Excludes group quarters. (B) - Base is less than 200,000 households. Individual outliers that highly influenced the mean value for debt categories were excluded. Federal surveys now give respondents the option of reporting more than one race. There are two basic ways of defining a race group. A group such as Black may be defined as those who reported Black and no other race (the race-alone or single-race concept) or as those who reported Black regardless of whether they also reported another race (the race alone-or-in-combination concept). This table shows data using the first approach (race-alone). The use of the single race population does not imply that it is the preferred method of presenting or analyzing data. The U.S. Census Bureau uses a variety of approaches. Because Hispanics may be any race, data in this table for Hispanics overlap slightly with data for the Black population. Data for American Indians and Alaska Natives are not shown because of their small sample size. The race or Hispanic origin of the householder designates the race or Hispanic origin of the household. The estimates in this table are based on responses from a sample of the population and may differ from the actual values because of sampling variability and other factors. As a result, apparent differences between the estimates for two or more groups may not be statistically significant. For information on sampling and nonsampling error see: <http://www.census.gov/programs-surveys/sipp/methodology/sampling.html>

Source: U.S. Census Bureau, Survey of Income and Program Participation, 2004 Panel, Wave 3

Internet Release Date: 8/21/2014



# Local Climatological Data (Providence, June 2020)

STATION	DATE	REPORT_TYF	SOURCE	AWND	BackupDirect	BackupDistal	BackupDistal	BackupElem	BackupEleva	BackupEleva	BackupEquip	BackupLatitu	BackupLongi	BackupName	CSD	CLDD	DSNW
7.2507E+10	6/1/20	0:51:00	FM-15	7	W	0.5	mi	TEMP, PRECIP, SNOW			PSY, SRG, SNOWBOARD			FAA CWO			
7.2						0.5	mi	TEMP, PRECIP, SNOW			PSY, SRG, SNOWBOARD			FAA CWO			
7.2						0.5	mi	TEMP, PRECIP, SNOW			PSY, SRG, SNOWBOARD			FAA CWO			
7.2						0.5	mi	TEMP, PRECIP, SNOW			PSY, SRG, SNOWBOARD			FAA CWO			
7.2						0.5	mi	TEMP, PRECIP, SNOW			PSY, SRG, SNOWBOARD			FAA CWO			
7.2						0.5	mi	TEMP, PRECIP, SNOW			PSY, SRG, SNOWBOARD			FAA CWO			
7.2						0.5	mi	TEMP, PRECIP, SNOW			PSY, SRG, SNOWBOARD			FAA CWO			
7.2						0.5	mi	TEMP, PRECIP, SNOW			PSY, SRG, SNOWBOARD			FAA CWO			
7.2507E+10	6/1/20	7:00:00	FM-12	4	W	0.5	mi	TEMP, PRECIP, SNOW			PSY, SRG, SNOWBOARD			FAA CWO			
7.2507E+10	6/1/20	7:51:00	FM-15	7	W	0.5	mi										
7.2507E+10	6/1/20	8:51:00	FM-15	7	W	0.5	mi										
7.2507E+10	6/1/20	9:51:00	FM-15	7	W	0.5	mi										
7.2507E+10	6/1/20	10:51:00	FM-15	7	W	0.5	mi										
7.2507E+10	6/1/20	11:51:00	FM-15	7	W	0.5	mi										
7.2507E+10	6/1/20	12:51:00	FM-15	7	W	0.5	mi										
7.2507E+10	6/1/20	13:00:00	FM-12	4	W	0.5	mi										
7.2507E+10	6/1/20	13:51:00	FM-15	7	W	0.5	mi										
7.2507E+10	6/1/20	14:51:00	FM-15	7	W	0.5	mi										
7.2507E+10	6/1/20	15:51:00	FM-15	7	W	0.5	mi										
7.2507E+10	6/1/20	16:51:00	FM-15	7	W	0.5	mi										
7.2507E+10	6/1/20	17:51:00	FM-15	7	W	0.5	mi										
7.2507E+10	6/1/20	18:51:00	FM-15	7	W	0.5	mi										
7.2507E+10	6/1/20	23:51:00	FM-15	7	W	0.5	mi										
7.2507E+10	6/1/20	23:59:00	SOD		W	0.5	mi										
7.2507E+10			2020-06-02T	FM-15	W	0.5	mi										
7.2507E+10			2020-06-02T	FM-12	W	0.5	mi										
7.2507E+10			2020-06-02T	FM-15	W	0.5	mi										
7.2507E+10			2020-06-02T	FM-15	W	0.5	mi										

What do you think is wrong with this data set?  
 What do you think is right with this data set?

## Local Climatological Data (LCD) Dataset Documentation

The Local Climatological Data (LCD) summaries provide a synopsis of climatic values for a single weather station over a specific month. The summaries are a product of surface observations from both manual and automated (AWOS, ASOS) stations with source data taken from the National Centers for Environmental Information's Integrated Surface Data (ISD) dataset. Geographic availability includes thousands of locations worldwide. Climatic values given include hourly, daily, and monthly measurements of temperature, dew point, humidity, winds, sky condition, weather type, atmospheric pressure and more.

**General conventions:** Times are in Local Standard Time (LST) unless otherwise noted. Temperatures are given in whole degrees F and tenths of degrees C (as noted below). Wind speeds are in miles per hour, with wind direction given using a 360 degree compass indicating the direction from which the wind was blowing with respect to true north (e.g. 360 = winds blowing from true north, 180 = south, etc.). Precipitation amounts are given in inches, usually to the nearest hundredth, unless otherwise noted. Trace amounts of precipitation are indicated with a "T." Departures from normal are given for some elements using 1981-2010 as the standard for normal. Further information concerning these generalities and units of measure are given where needed for each section below.

### Table of Contents:

- Daily Summary/Monthly Summary
- Hourly Observations

Go to [menti.com](https://www.menti.com) and type in the code or follow  
this link

<https://www.menti.com/c21osopcv5>

**\*\*\*\*\*Mentimeter questions\*\*\*\*\* Link: 30 64 41**



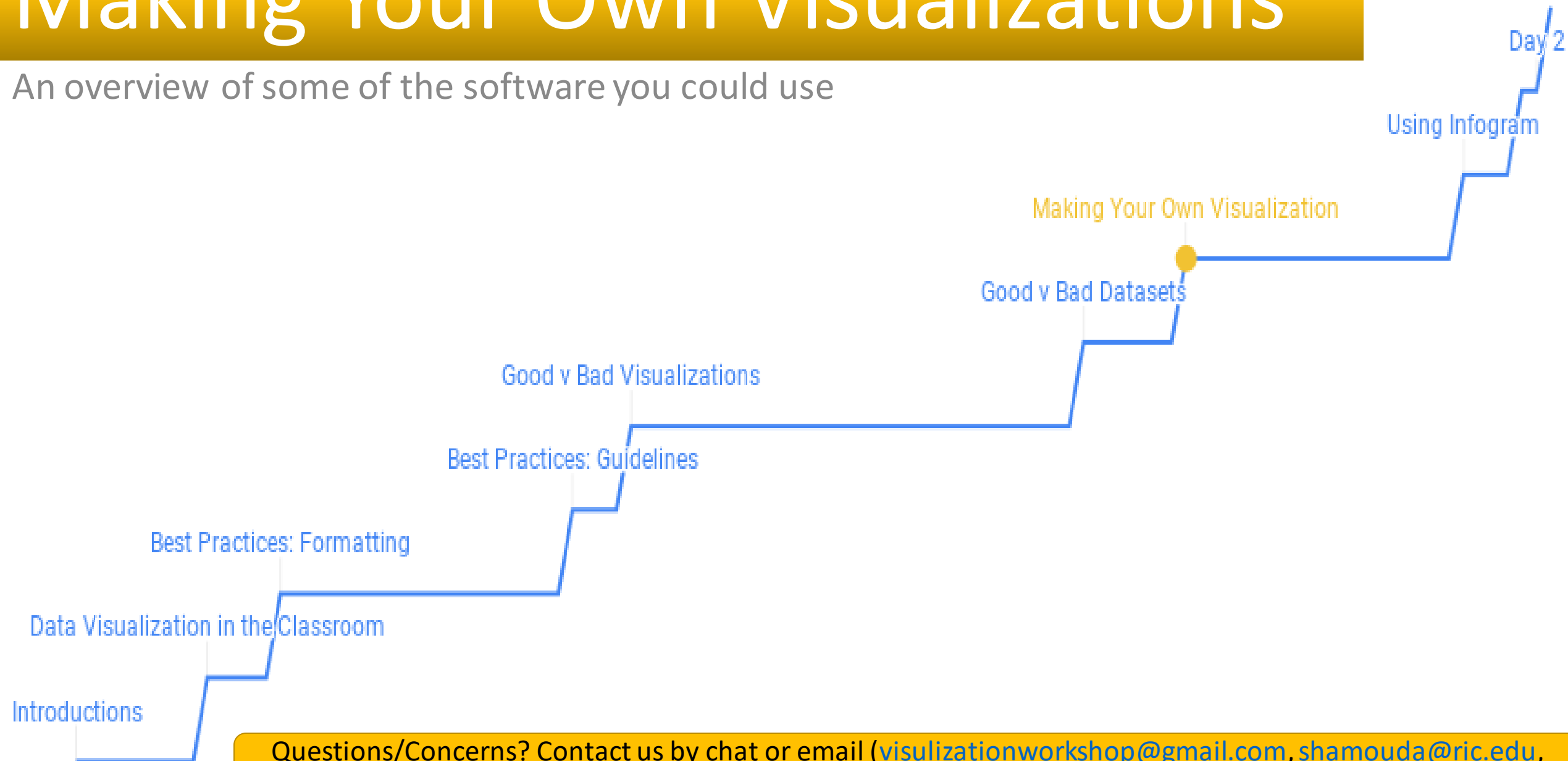
# Local Climatological Data (Providence, June 2020)

- Most of the column names are easy to decipher
- Metadata is reasonably presented

- There are some blank rows with acronym headers
- You have to split the cells to separate date and time
- Not all of the metadata you need or would like to look at is in one place

# Making Your Own Visualizations

An overview of some of the software you could use



Questions/Concerns? Contact us by chat or email ([visulizationworkshop@gmail.com](mailto:visulizationworkshop@gmail.com), [shamouda@ric.edu](mailto:shamouda@ric.edu), [lcendella\\_2763@email.ric.edu](mailto:lcendella_2763@email.ric.edu))

# Excel and Google Sheets

- These are probably the most common programs you will use
  - The following programs even accept uploads from them
- Google Sheets is free and allows for easier collaboration
  - It's user-friendly
- Excel is great with big data, but generally requires training on how to utilize its full capabilities
  - Office 365 does allow collaboration between users now

# Power BI

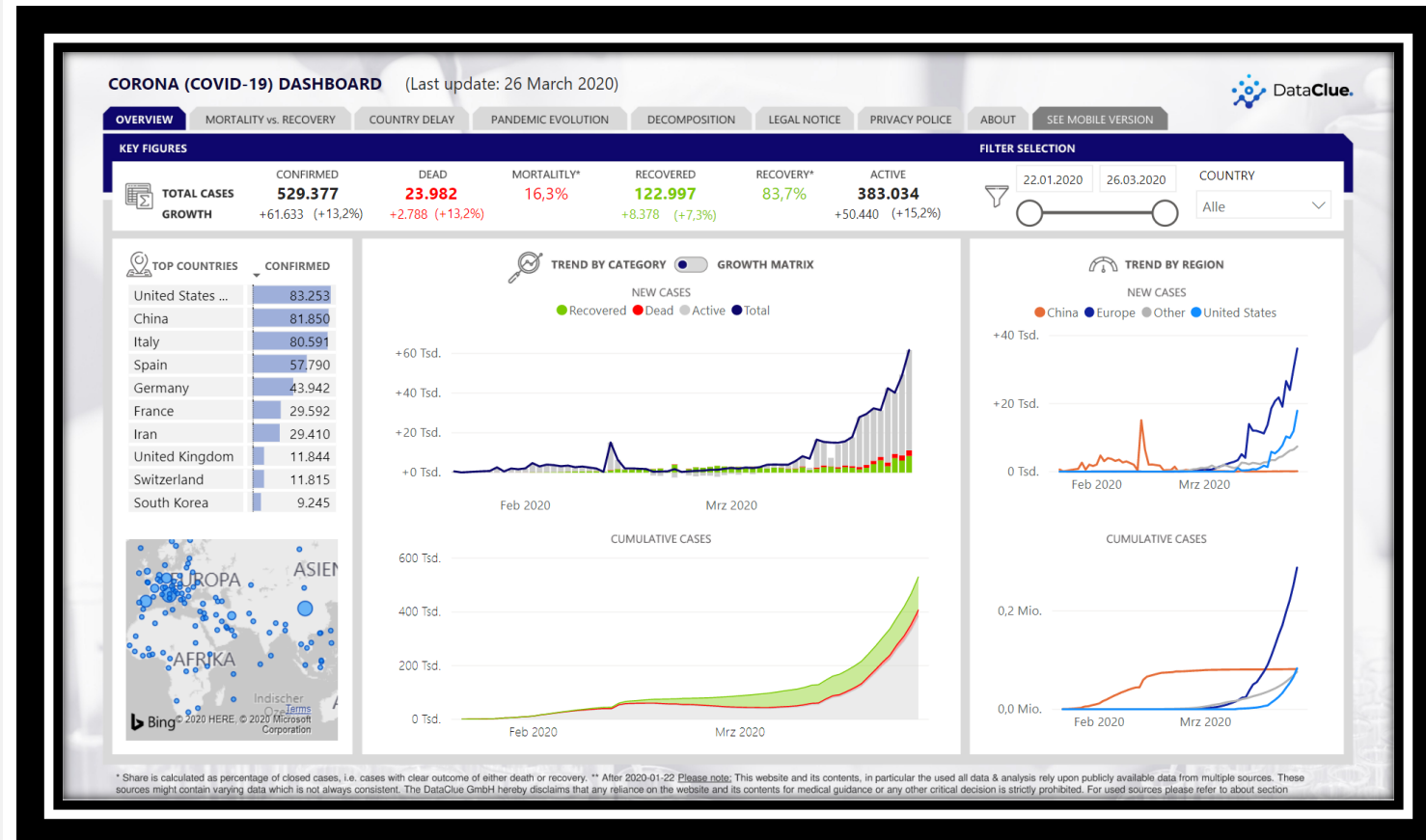


- A Microsoft product for businesses to make interactive data visualizations
- Has guided learning and a free 4-week course
  - Also has a free downloadable version where you can make Power BI reports or publish them on their website
- Requires a Windows operating system; to collaborate on one project you need to pay for a Power BI Pro license
- Only 501c organizations receive discounted/donated software; schools with at least 1,000 possible student users can apply for the Enrollment for Education Solutions (EES)

# What can it do?



- Power BI says its strength is making complex data more comprehensible, which leads to data-driven decision-making
- Power BI can create all types of visualizations using a drag-and-drop method
- Let's take a look at Power BI's featured Dashboard (how visualizations are presented to others)



# How can I get it?



<https://www.microsoft.com/en-us/download/details.aspx?id=58494>

Link to free download

## Choose Power BI Pro

- Self-service and modern BI in the cloud
- Collaboration, publishing, sharing, and ad-hoc analysis
- Fully managed by Microsoft

### Power BI Pro

**\$9.99**

Monthly price per user

[Buy now >](#)

[Try free >](#)

## Add Power BI Premium

- Enterprise BI, big data analytics, cloud and on-premises reporting
- Advanced administration and deployment controls
- Dedicated cloud compute and storage resources
- Allows any user to consume Power BI content

### Power BI Premium

**\$4,995**

Monthly price per dedicated cloud compute and storage resource with annual subscription

[Request a consultation >](#)

Pricing is simple and intuitive!

# Tableau

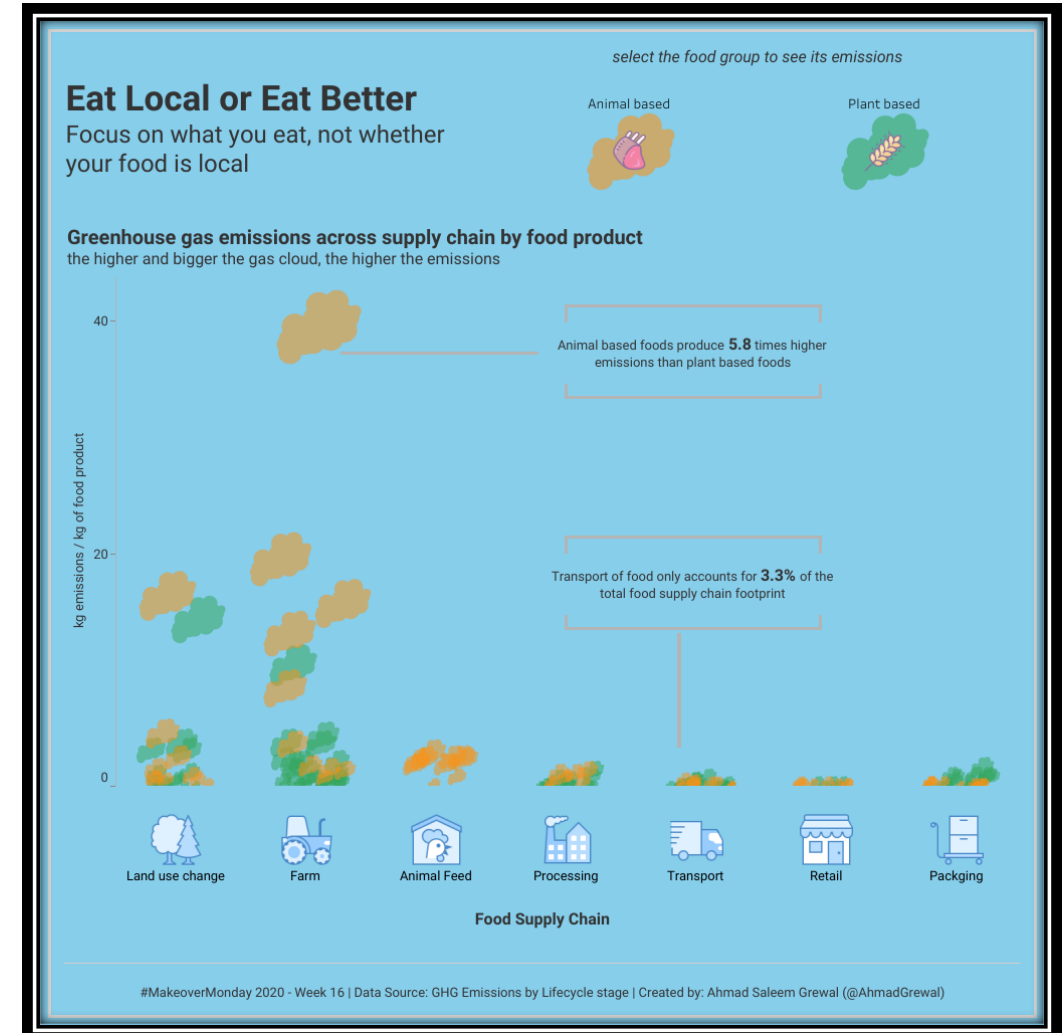


- Designed for professional use, has hours and hours of training videos
- Once verified, teachers and students get free copies of the software to use
  - The instructor resource page even has ready-made course materials, ways to contact other teachers using Tableau, a free guidebook for students, and a knowledge base for FAQs
- You can even get travel grants to go to academic conferences!

# What can it do?



- Mission Statement: We help people see and understand data
  - Very true for people that see the end product!
- Learning to use Tableau or Power BI is no light undertaking
  - Tableau can handle big data for big businesses
- It will take a long while to be able to utilize either program to the most of its capabilities
- We can take a look at Tableau's featured content here





# How can I get it?



<https://www.tableau.com/academic/teaching/course-licenses>

Link to request free software

## Tableau Creator

\$70

USD/user/month  
billed annually

Discover insights faster by combining a powerful suite of products for visual data prep, best-in-class analytics, and secure collaboration to support your end-to-end analytical workflow.

Buying for one person has  
just the one option

Buying for a team has  
several options available

DEPLOY WITH TABLEAU SERVER  
On-premises or Public cloud

DEPLOY WITH TABLEAU ONLINE  
Fully hosted by Tableau

### Tableau Creator

Every deployment requires at least one Creator

\$70

user/month | billed annually

Discover insights with a powerful suite of products that support your end-to-end analytics workflow.

### Tableau Explorer

\$35

user/month | billed annually  
min. 5 Explorers required

Explore trusted data and answer your own questions faster with full self-service analytics.

### Tableau Viewer

\$12

user/month | billed annually  
min. 100 Viewers required

View and interact with dashboards and visualizations in a secure, easy-to-use platform.

# Infogram

infogr.am

- A way to make infographics
- You can focus solely on the visualization side of it, but there are many options to add different graphics or animations
- The free option limits you to 10 projects, max 5 pages each
  - You also have no option to download the project; it becomes public and you can get a shareable link to view it
  - You can take a nice screenshot of your project from the download page, though

*The Great Gatsby: Chapter 1 Word  
and Theme Frequency*

Daisy

Desires

Assumptions

Gatsby

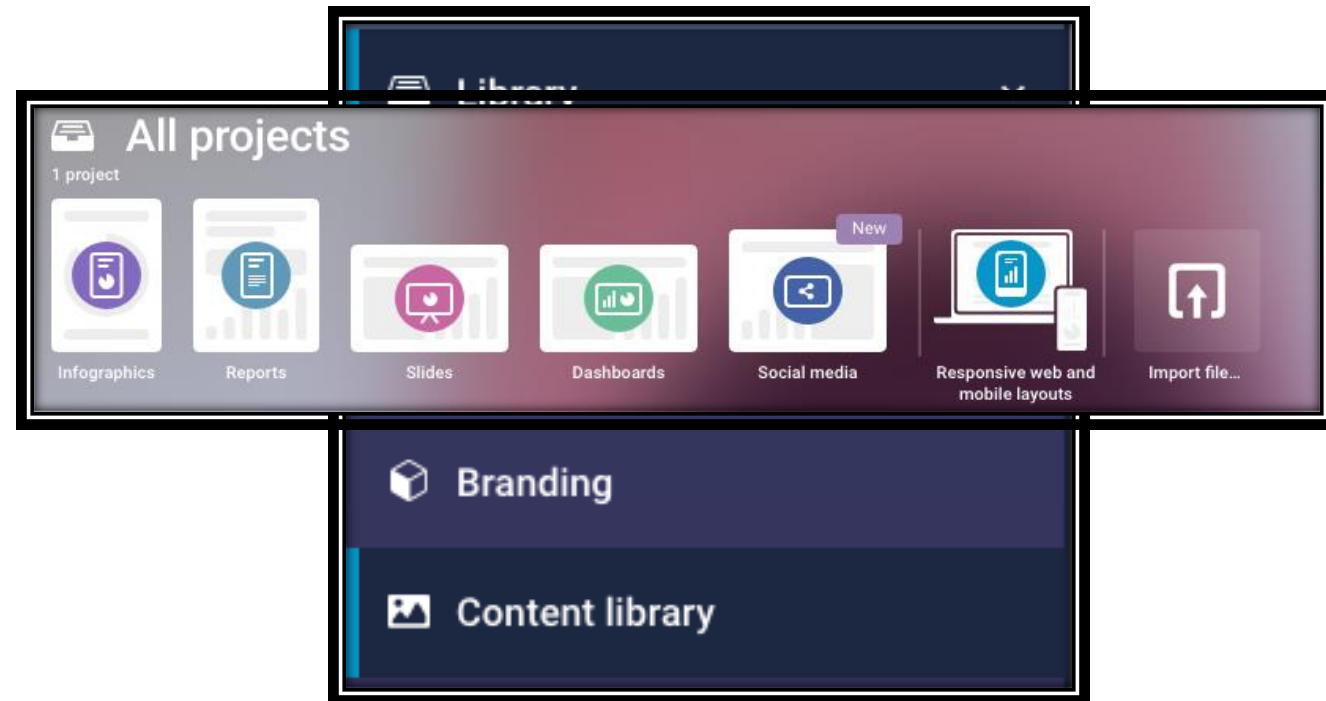
Society

# How does it work?

infogr.am

## The Dashboard – New Project Options

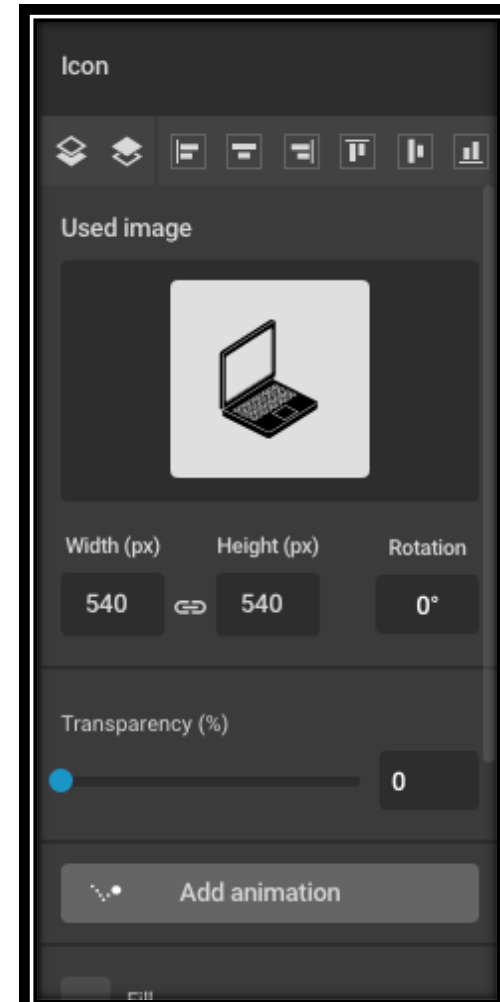
- Choosing one takes you to the relevant template section of their library
- From there you can search for a template, you look at the other categories' options
  - Content Library is every icon they have and what you uploaded



# How does it work?

## The Project Toolbar and Element Settings

- There are so many options once you start looking!
  - You have Text, Charts, Maps, Elements, Graphics, Shapes, and Integration (video, links, data)
- Anything on the page can have its settings changed once you click on it
  - Order, color, size, transparency, and more



# How does it work?

The logo for infogr.am is a red speech bubble shape with the text "infogr.am" in white lowercase letters inside.

## Making a Visualization

- Look at your data and choose a chart type that works, not one that looks interesting
- Let's say I want to make a poster for the Travel Abroad club at school
  - I want to show the cost per person to different destinations the group may go to
- What chart type do you think would work for this data?

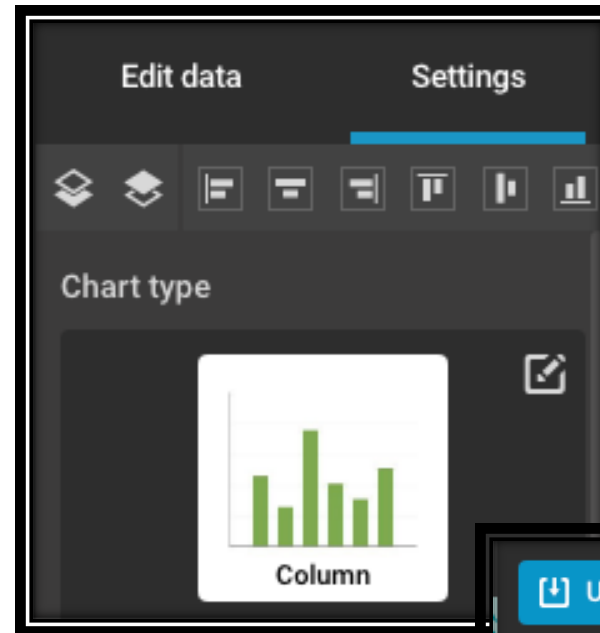
Destination	Cost
Paris	5,000
Egypt	4,500
Tokyo	8,600
London	3,000
Brazil	3,300

# How does it work?

infogr.am

## Importing Data

- Since the data set is so small, I can just edit the chart element directly
  - While the bar chart is selected, click "Edit data" to bring up the spreadsheet and fill in your data
- What if I already have a spreadsheet of data? Simple!
  - The "Edit data" tab allows you to upload an Excel file, a .csv file, a Google Sheets file, and much more
- With .csv, make sure to click on the little gear icon to change how the data is separated



I chose to use a bar chart

The screenshot shows a spreadsheet editor with a dropdown menu open. The spreadsheet has columns A and B, and rows 1 through 5. The data in column A is 'Paris', 'Egypt', 'Tokyo', and 'London'. The data in column B is 'Cost', '5,000', '4,500', '8,600', and '3,000'. The dropdown menu is open over the 'Cost' cell, showing options: 'Auto', 'Period (.)', 'Comma (,)', and 'Aggregate data'. The 'Period (.)' option is selected.

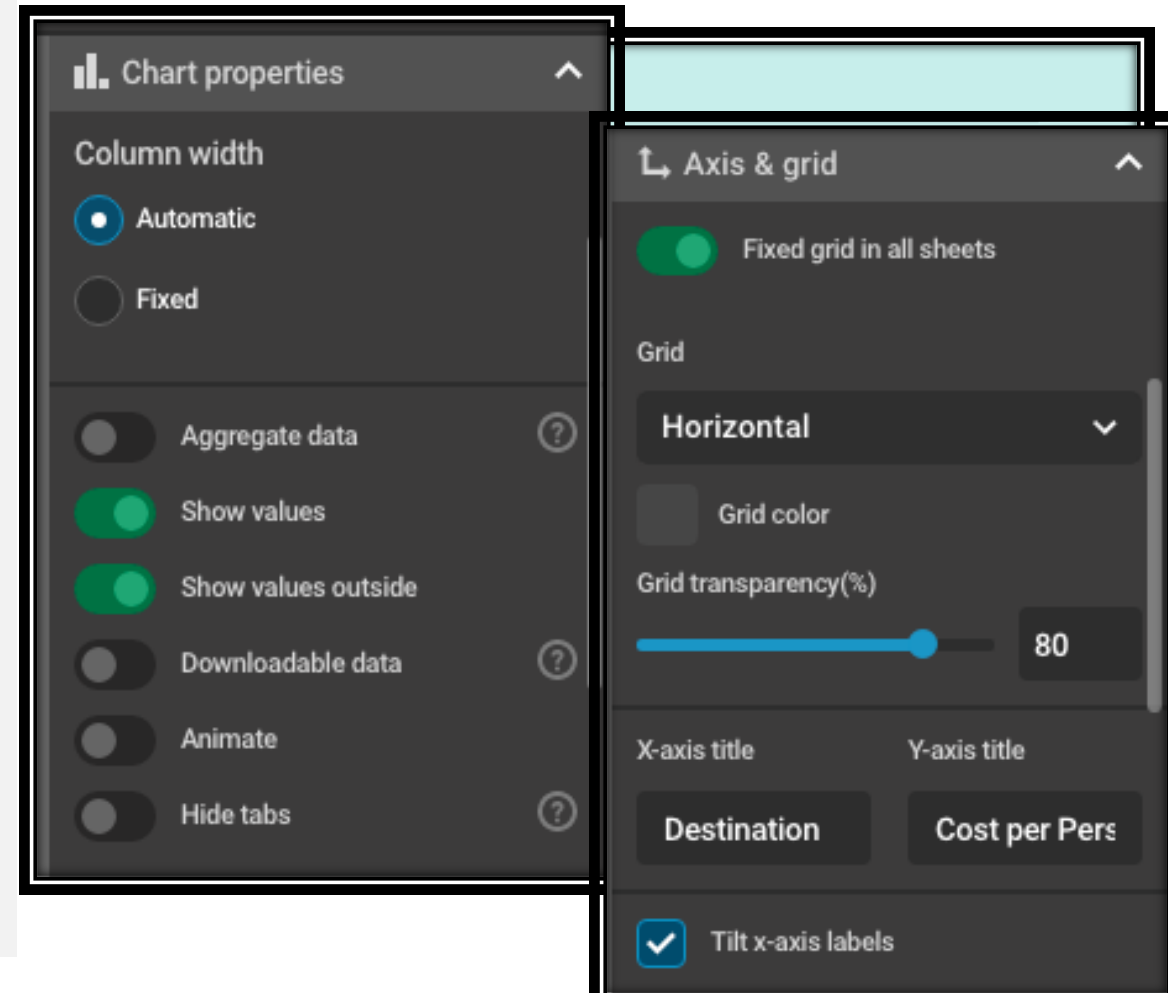
	A	B
1		Cost
2	Paris	5,000
3	Egypt	4,500
4	Tokyo	8,600
5	London	3,000

# How does it work?

infogr.am

## Finalizing the Graph

- Going back to the settings, I want to add some elements
- I want data labels since this will be a printed poster
  - Go to "Chart properties" --> "Show values" and/or "Show values outside"
  - If I share the poster from Infogram digitally, I don't need data labels since the labels appear when you hover over a column
- I want axis titles
  - Go to "Axis & grid" and scroll down to find the text boxes for each axis
- There are many more options to play around with if you make a chart on Infogram



# Piktochart



- A way to make infographics, flyers, presentations, reports, or posters
- For academic use, Individual PRO plans cost \$40 per person and PRO Team plans cost \$200 a year
  - You need the team plan to allow for design collaboration
- The free version lets you make 5 visualizations
- You can download your work as a PNG (free version), or as a PDF or PowerPoint (paid versions)
  - You can also share your visualization publicly for free, and get a link for people to view it





## COURSE EXPECTATIONS



## OFFICE HOURS



## CLASSROOM PHILOSOPHY

Come prepared, leave with a deeper understanding of how the human body works.

## HOW TO SUCCEED IN THIS COURSE

Keep up with the lecture material.



Study in group and explain hard concepts to others to check for understanding.



Ask the professor about learning strategies specific for this class



Study at least 10 minutes everyday.



Attend office hours, open labs and review sessions



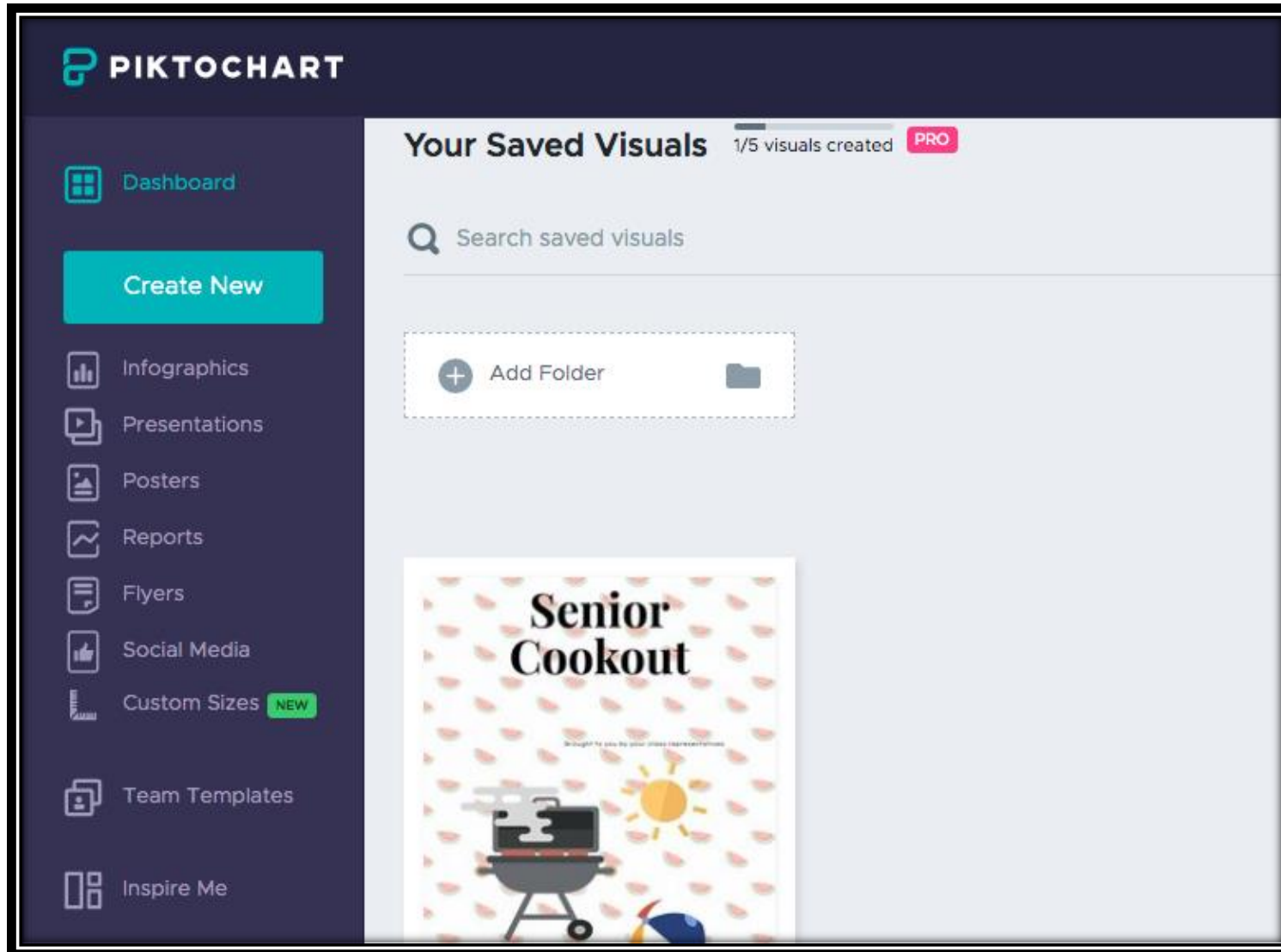
Make use of the Bio tutoring



## Dr. Maia's Syllabus

- A different approach
- More engaging than a traditional syllabus layout

# How does it work?



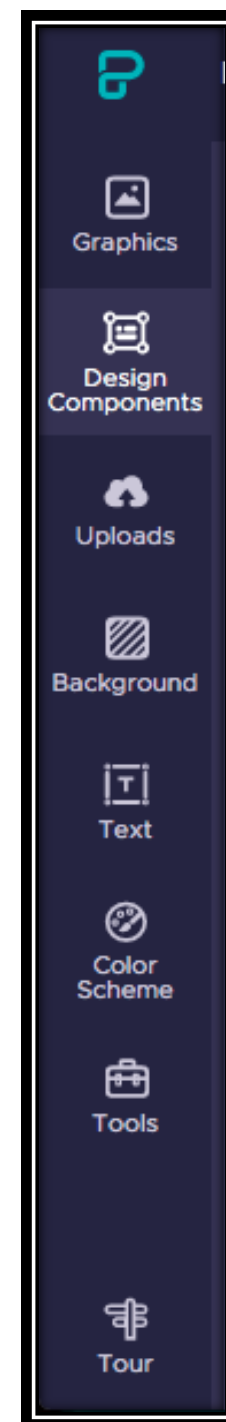
## The Dashboard

- Shows your previous work
- List of options on the side brings up relevant templates
- Create New brings up the list of options with a short description for each one

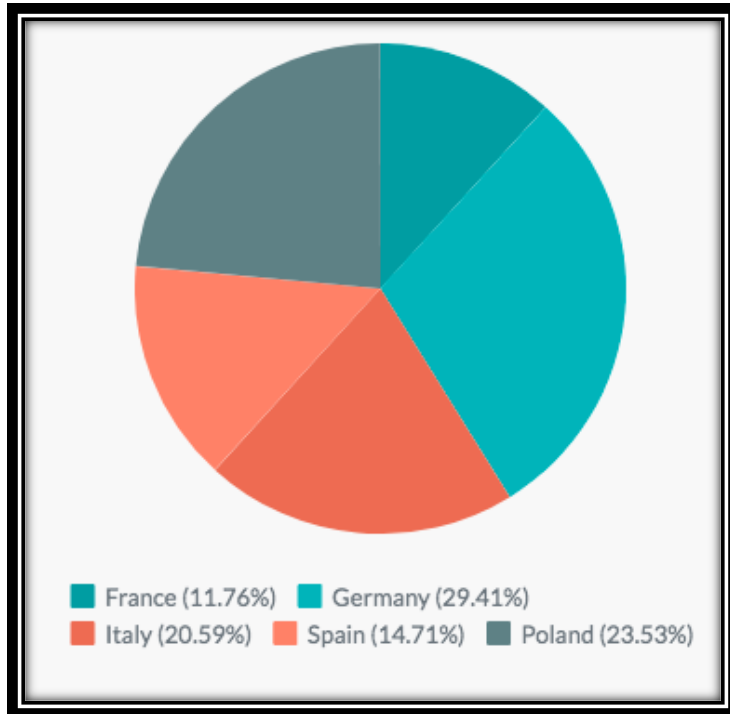
# How does it work?

## The Toolbar

- Pretty much every tool is self-explanatory
- Graphics has icons, lines, photos, and other stylistic elements
- Design Components allows for the specific block types that essentially make up each project
  - Lists, Timelines, and Comparisons
- Tools has your Charts, Maps, and Video options
  - For charts, you can edit the data in the software, upload your data, or make dynamic data by inserting the Google Sheets link
  - Watch out for some of these charts! They do not follow the best practices

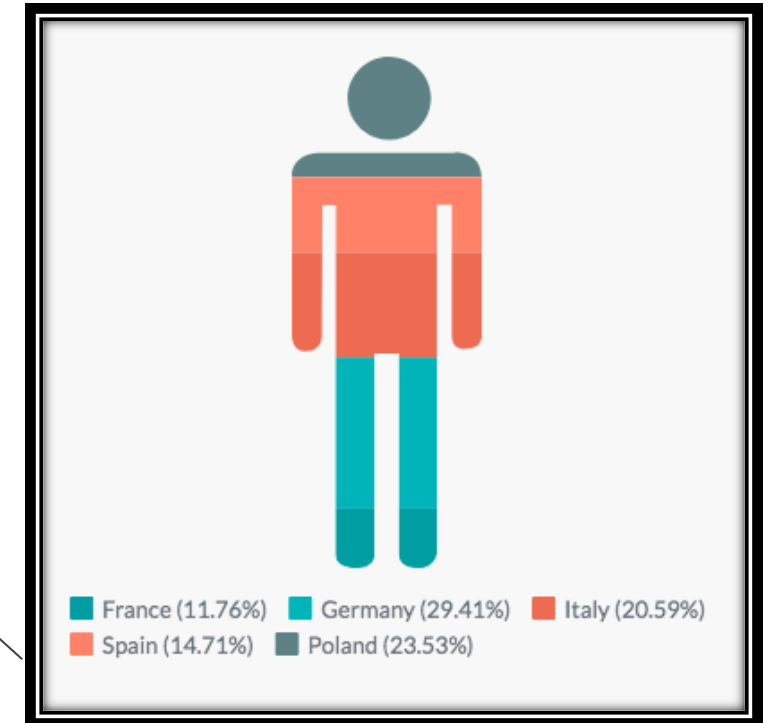


# What to look out for



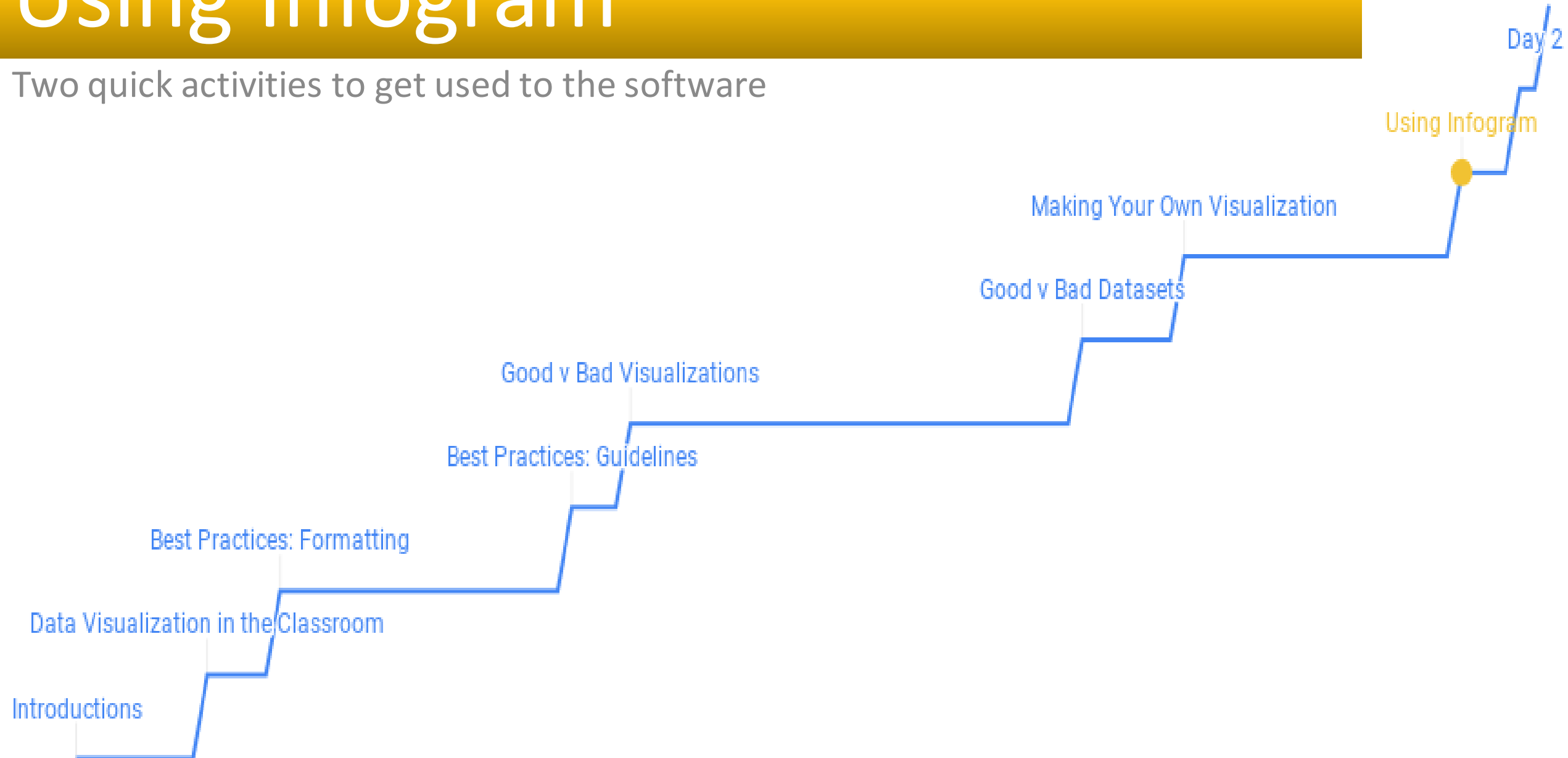
Incorrect ordering on pie charts (you will have to order it yourself in the data)

Odd shapes lead to confusion



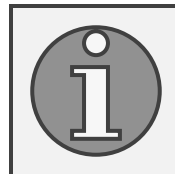
# Using Infogram

Two quick activities to get used to the software



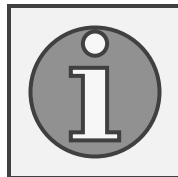
# Course Descriptions

- With Infogram, you can make class selection more engaging for incoming or returning students
  - For example, I made a course description
- I did this by:
  - Going to Infogram and made an account (signed in with Google)
  - Made a new project with the "Long Resume" template
  - Filled in, rearranged, deleted, and changed what I needed



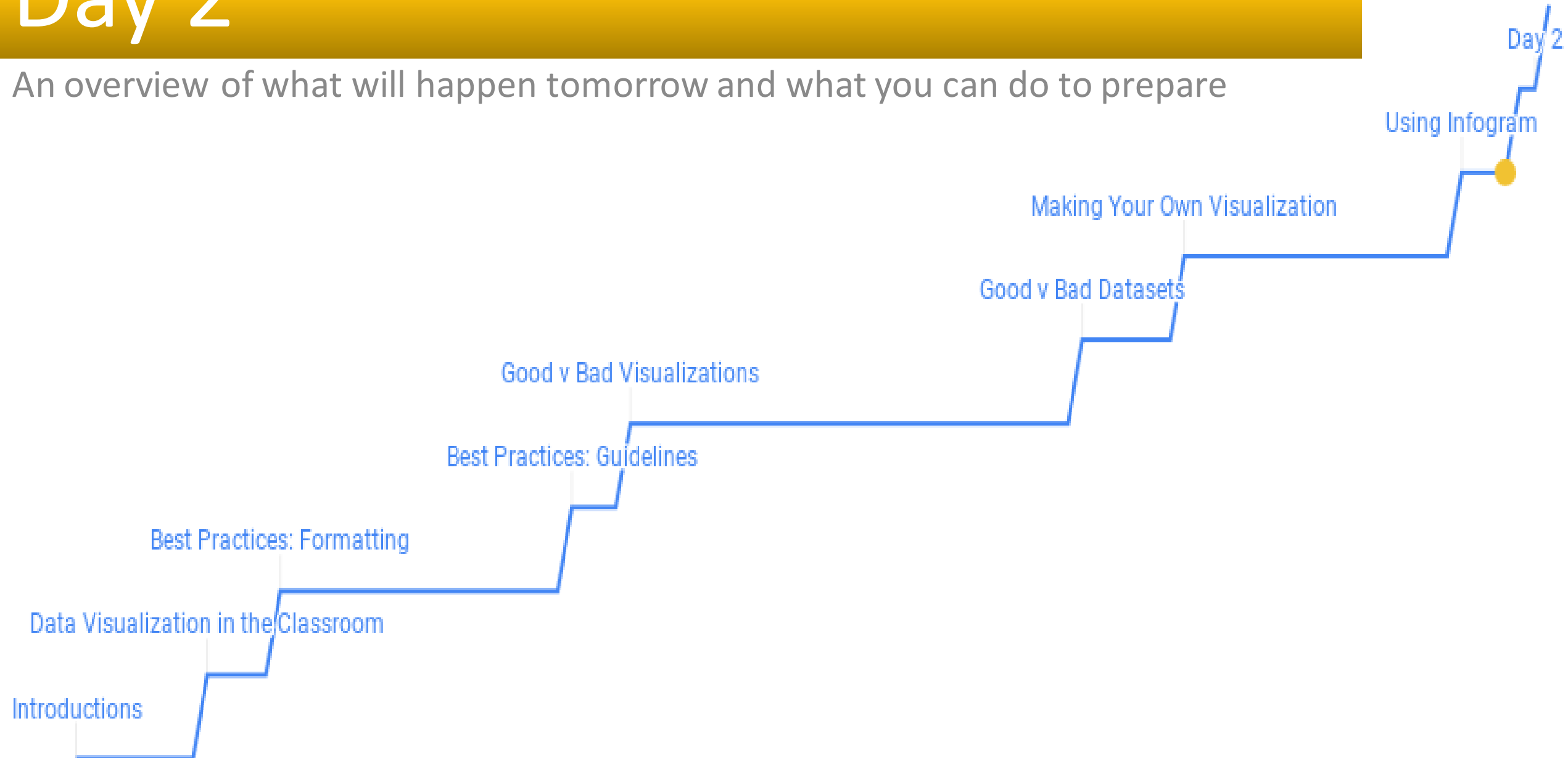
# Student Reports

- Say you want parents and students to easily track their attributes in a class (attendance, completed work, etc.)
  - For example, I made a student report
- With the paid version you can link a Google Sheets spreadsheet to the table for live data
  - Any changes you make to the spreadsheet show up in the report
  - Any parents or students that have the link to the report can see changes in real time



# Day 2

An overview of what will happen tomorrow and what you can do to prepare





# Creating a course activity

- Working in groups, make a data visualization activity that you could integrate into your curriculum
  - We will be here to supervise and answer questions!
- In the meantime, think of:
  - What tool you would like to use
  - What you want the data story to be
  - Where you will get the data
  - What the end product should be (physical or digital visualization)
- There is a document of these questions in the Google Drive
- If you have any questions, feel free to contact us
  - We have a collection of data bases for you to look at if you cannot find one

# End of Day 1

**Thank you!**  
**Any Questions?**

