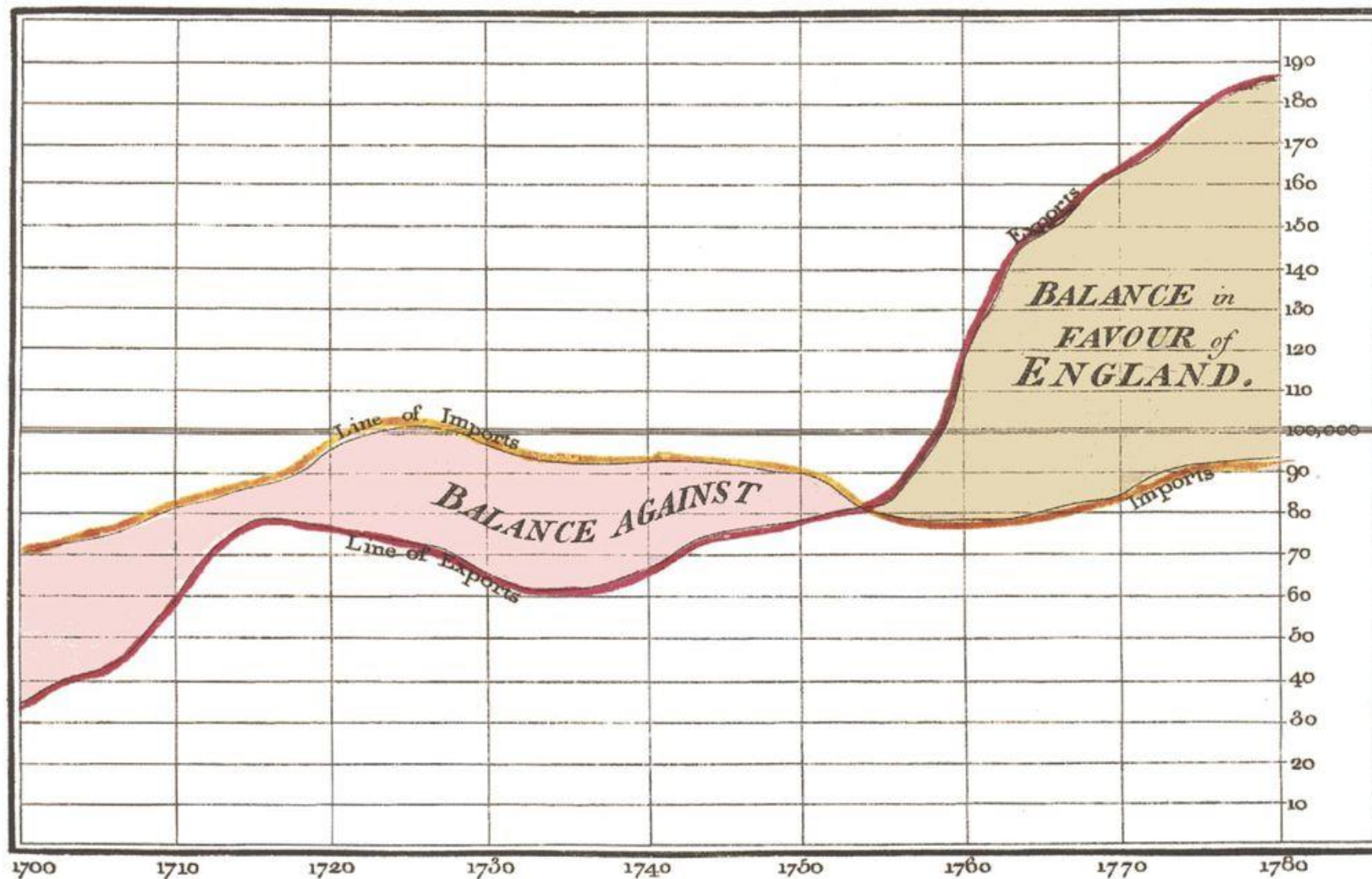


INTRO TO DATA VISUALIZATION: SPRING INTO RESEARCH 2021

Katie Kuiper & Keiko Bridwell
katherine.kuiper25@uga.edu
keiko.bridwell@uga.edu



Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



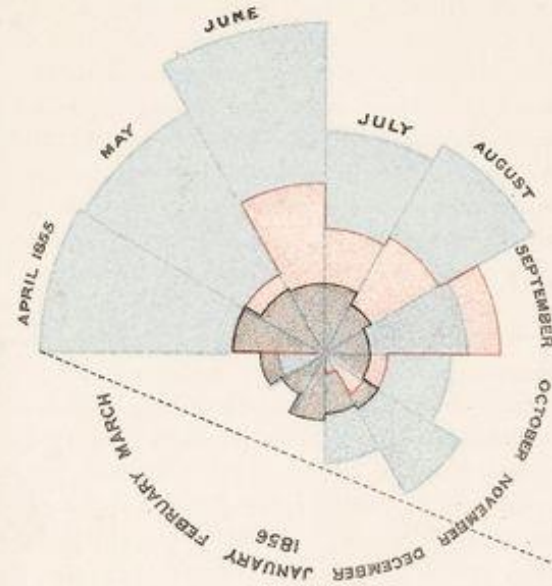
The Bottom line is divided into Years, the Right hand line into £10,000 each.

Published as the Act directs, 1st May 1786, by W^m Playfair.

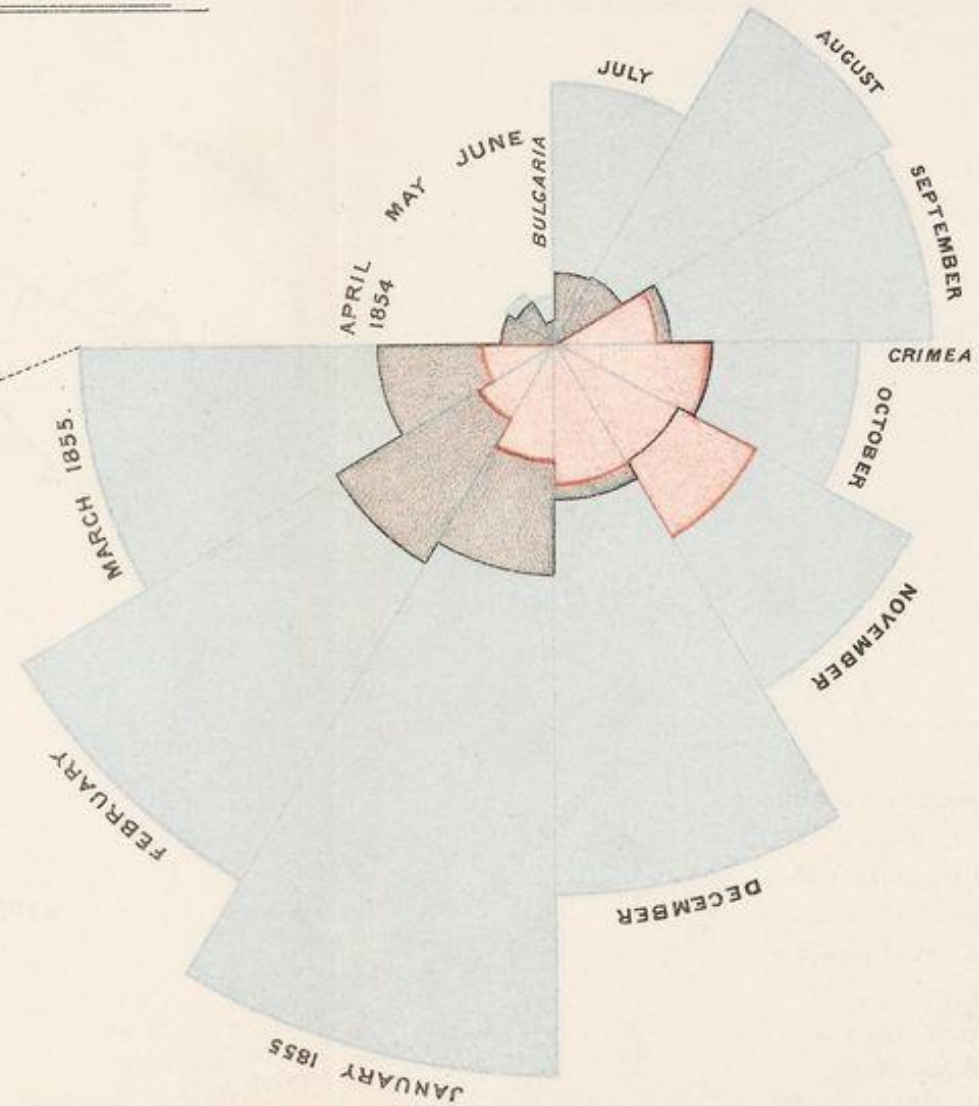
Needle sculpt 352, Strand, London.

DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.

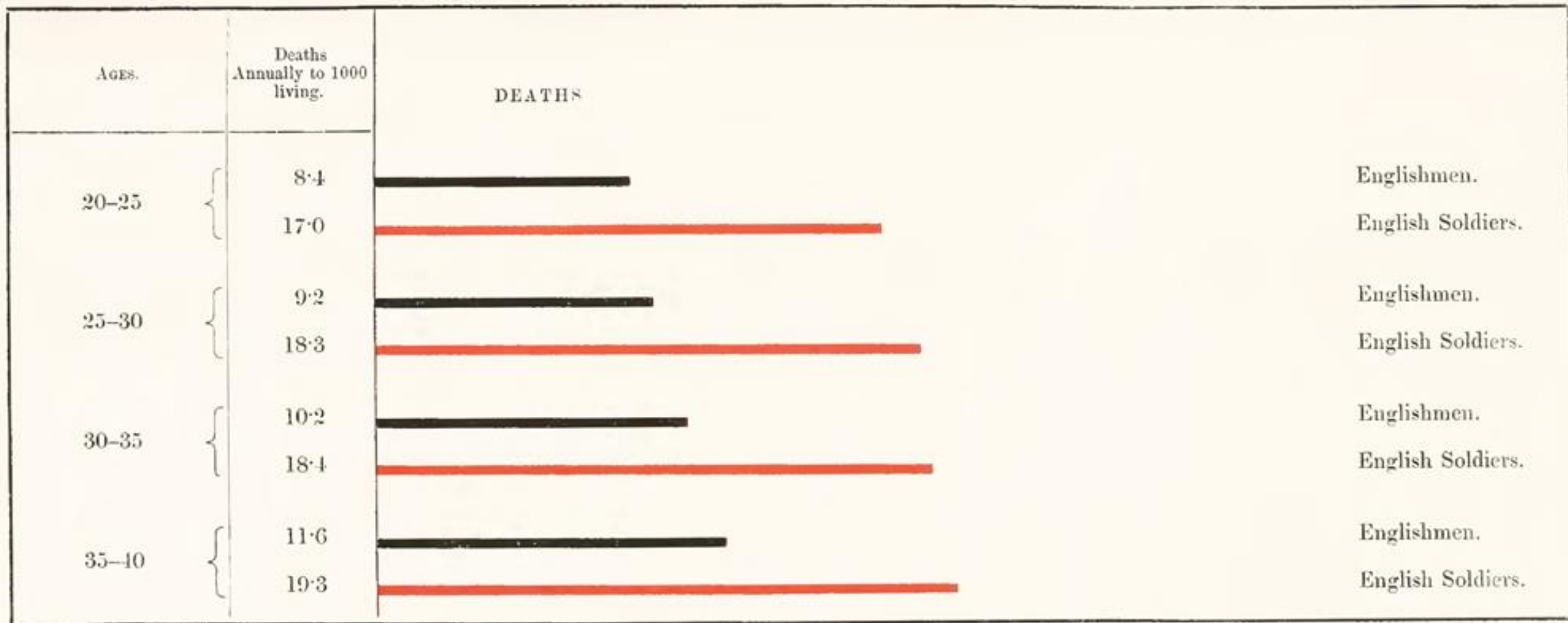


1.
APRIL 1854 TO MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.
The black line across the red triangle in Nov^r 1854 marks the boundary of the deaths from all other causes during the month.
In October 1854, & April 1855, the black area coincides with the red, in January & February 1855, the blue coincides with the black.
The entire areas may be compared by following the blue, the red & the black lines enclosing them.

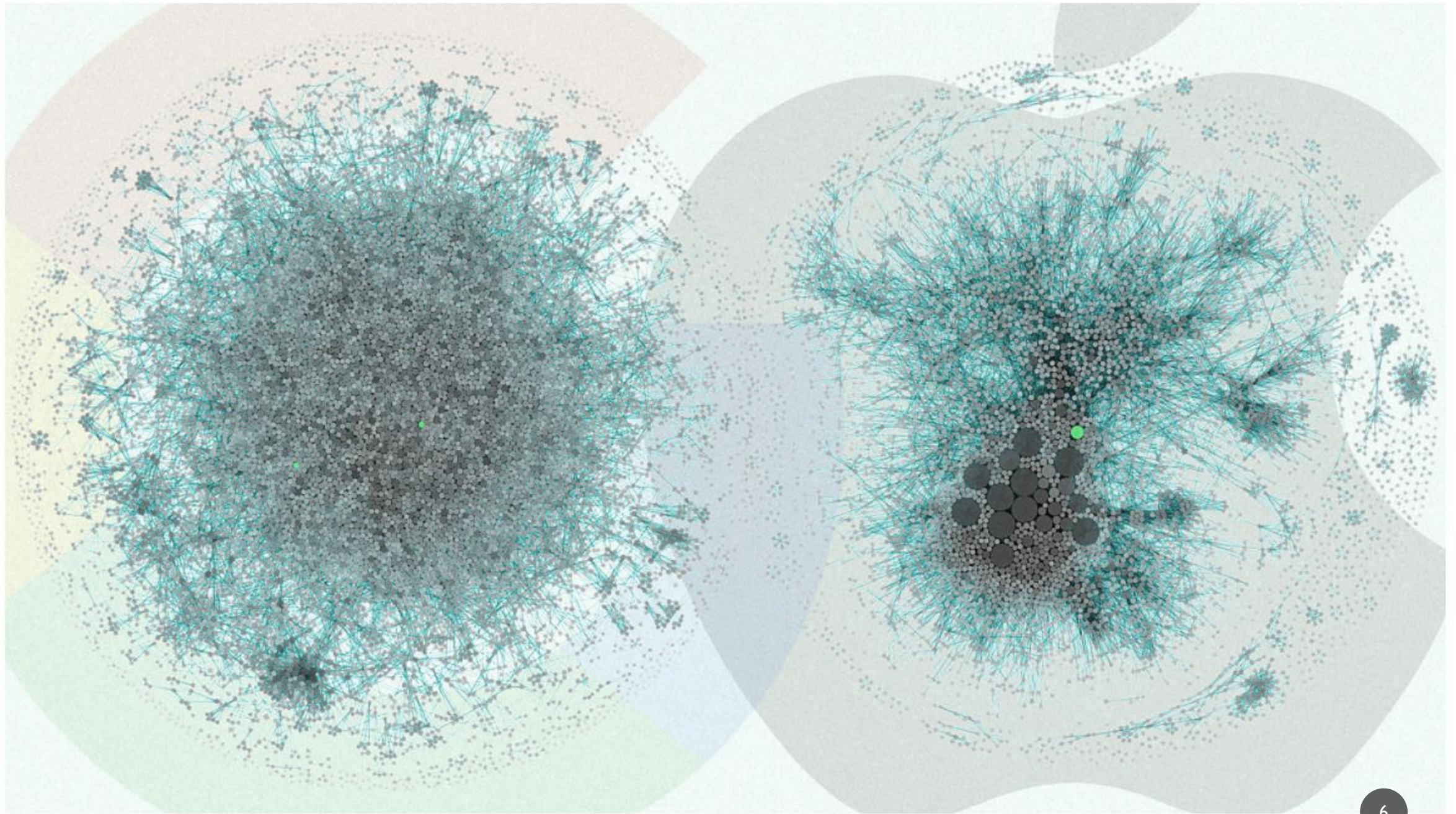
Representing the Relative Mortality of the Army at Home and of the English Male Population at corresponding Ages.



JAMES LEWIS, del.

VISUALIZATION
TRANSFORMS DATA
INTO IMAGES THAT
EFFECTIVELY AND
ACCURATELY
REPRESENT
INFORMATION
ABOUT THE DATA.

Schroeder et al. 1998: The Visualization
Toolkit



MAIN GOALS OF DATA VISUALIZATION



1. EXPLORE AND
UNDERSTAND



2. TO ANALYZE



3. PRESENT AND
COMMUNICATE
RESULTS



4. PROMOTE
ENGAGEMENT

COMMON TERMS

- Statistical graphics
- Infographics
- Information visualization
- Data visualization

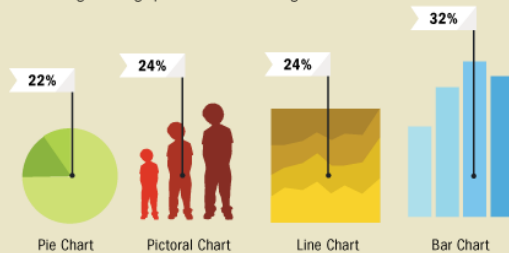
INFOGRAPHIC OF INFOGRAPHICS

Data visualization is a popular new way of sharing research. Here is a look at some of the visual devices, informational elements, and general trends found in the modern day infographic.

DESIGN

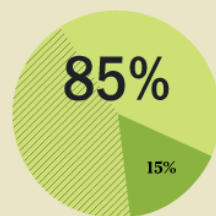
CHART STYLE

Percentage of infographics with the following charts:



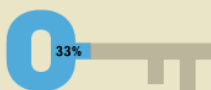
FONT

■ Sans Serif
 ▨ Condensed Sans Serif
 ■ Serif



KEY INFO

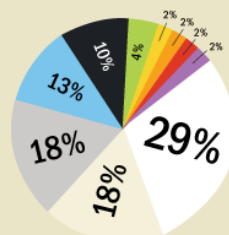
Percentage of infographics with key:



Average number of symbols per key: **5.1**

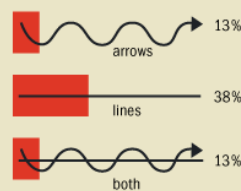


BASE COLOR



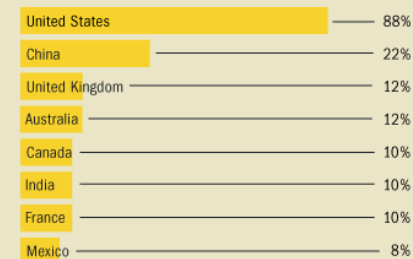
NAVIGATIONAL ICONOGRAPHY

Frequency of arrows & connecting lines in infographics:



CONTENT

COUNTRIES FEATURED

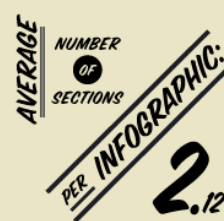


THEME

Relative popularity of different infographic themes:



SECTIONS



CREDITED SOURCES

Average number of sources per infographic: **2.29**



TITLE

Average number of words per infographic title: **4.36**

“RICHEST AND POOREST AMERICAN NEIGH

CONCEPT & DESIGN Ivan Cash

SOURCE 49 infographics collected at random from www.good.is/infographics

BEFORE GETTING STARTED

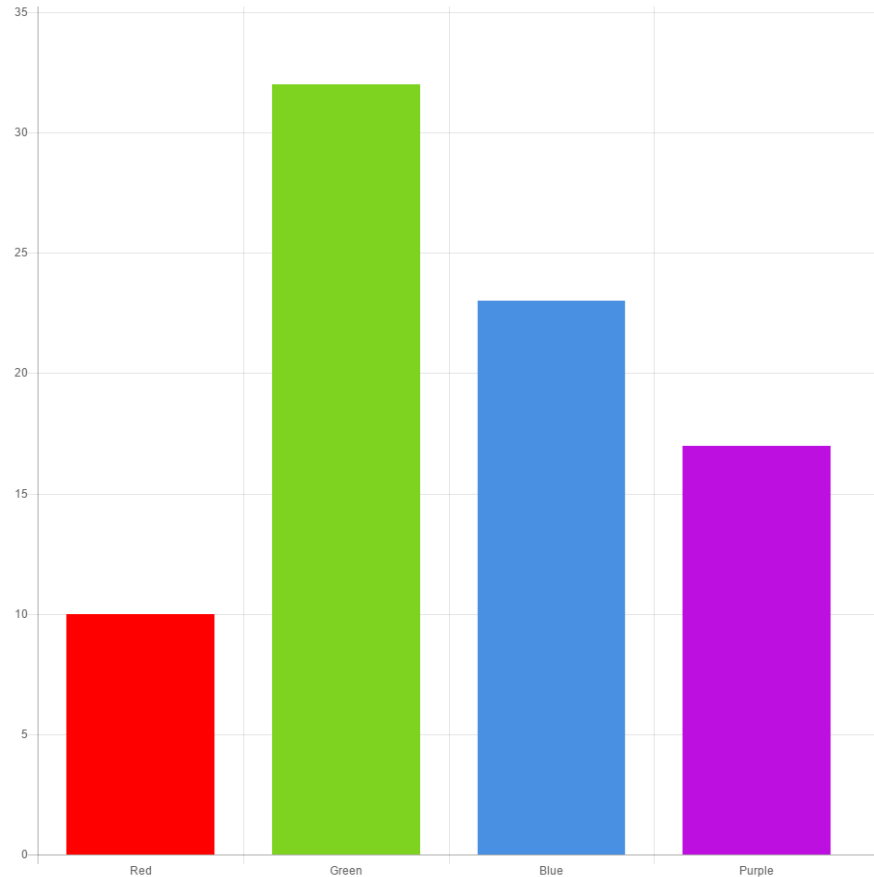
- In order to select the best visualization for your data, ask yourself:
- Who is my audience?
- What am I trying to convey and accomplish?
- What kind of data am I using?

AUDIENCE AND GOALS

- Who is the visualization for?
- Is the goal exploration, explanation, or persuasion?
- Depending on these, decisions on your visualizations might differ.
- You might want to make your visualizations static,
- or dynamic!

DATA TYPES

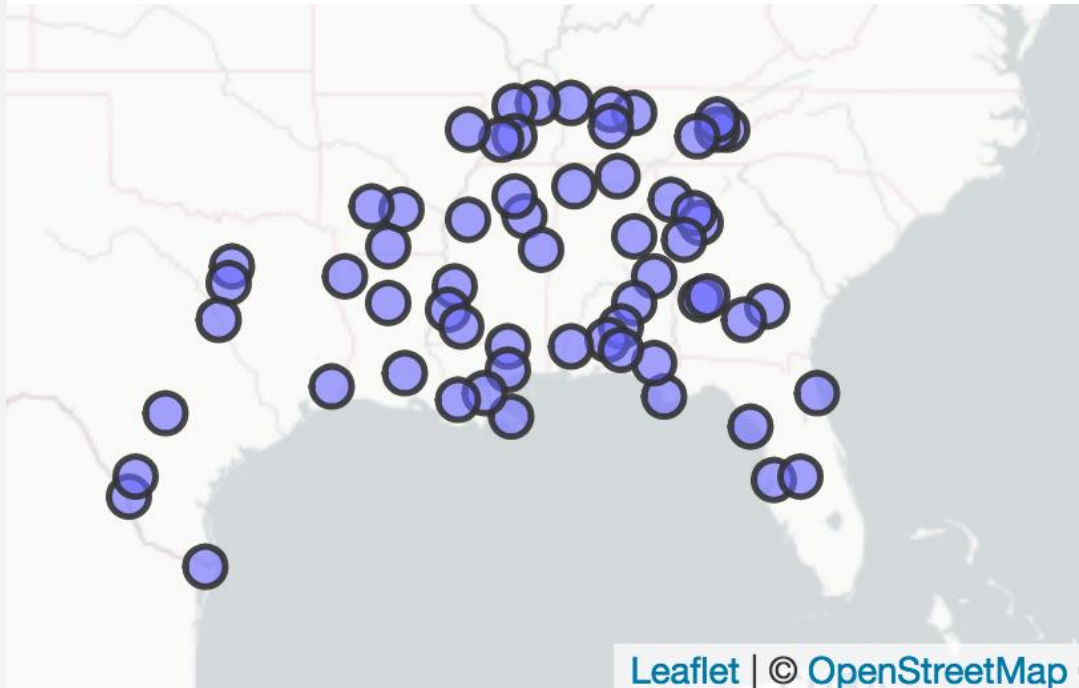
- **Quantitative:** numerical values; data that is measured or counted
 - **Discrete-** countable data
 - population, money, word frequency
 - **Continuous** – data that can take any value temperature, length, time
 - temperature, length, time
- **Qualitative/categorical:** data that falls into labeled categories that occur in no particular order
 - college major, text genre, demographic categories



DATA PATTERNS

- How do your data relate to each other?
- Common data relationships include:
 - **Part-to-whole:** ratio in comparison to the whole; ex: bar charts, pie charts
 - **Distribution:** an illustration of how the data are spread/located, along with where they tend to center, outliers, etc. ex: boxplots
 - **Correlation:** a comparison of how two or more variables may have a positive or negative relationship (ex: scatterplots)

DATA PATTERNS



- **Time series-** showing at least one variable over a period of time, usually displaying changes over time.
- **Geospatial:** comparison of a variable across a map or specific layout.
 - Ex: This map, showing the speakers in the Digital Archive of Southern Speech (Kretzschmar et al. 2012)



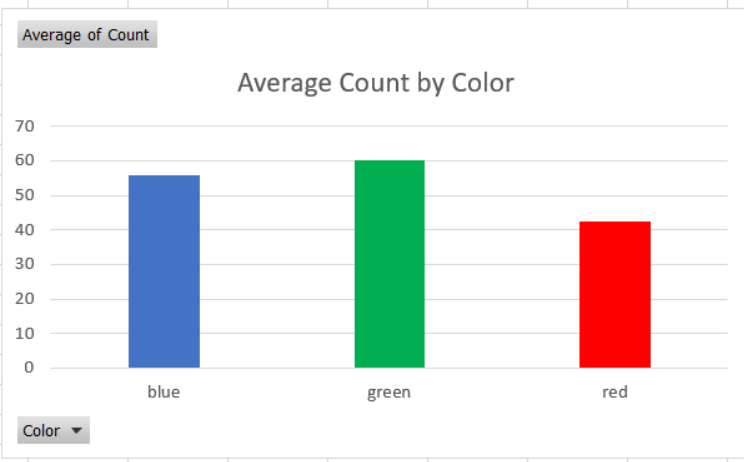
COMMON TYPES OF VISUALS

- Charts
- Tables
- Graphs
- Maps
- Infographics
- Dashboards

- As a spreadsheet software, stores and displays data in a tabular format: data are stored in “cells” and can be manipulated via formulas
- Provides easy and effective comparisons
- Useful for statistical operations and analysis of data
- Does not require scripting knowledge (but has the option to use that)



	A	B	C	D	E	F	G	H	I	J	K
1	Color	Count									
2	red	75									
3	blue	8									
4	red	13									
5	green	41									
6	blue	65									
7	blue	100									
8	green	99									
9	red	62									
10	green	51									
11	red	46									
12	blue	27									
13	red	27									
14	blue	90									
15	green	81									
16	blue	55									
17	blue	45									
18	red	31									
19	green	16									
20	green	72									
21											
22											
23											
24											
25											
26											
27											



Insert Chart

Recommended Charts All Charts

Recent Templates Column Line Pie Bar Area X Y (Scatter) Map Stock Surface Radar Treemap Sunburst Histogram Box & Whisker Waterfall Funnel Combo

Clustered Column

Chart Title

Count

OK Cancel

- Includes a wide range of different chart options:
 - Column chart
 - Histogram
 - Scatterplot
 - Box and whisker chart
 - Area chart





- Extremely popular worldwide and widely used in business and public sectors
- Features powerful business intelligence tools
- A bit steeper learning curve than Excel
- Includes different data processing options and analytic tools
- Optional plug-ins to other tools to house data and visualizations
- Tableau Public is free!



Data Analytics

hw.compare

Search

Tables

City Born
City Home
Hnr1
Hnr3
Hw Perc Intensity
ID
Occupation1
Participant
Passage ID
Prev Seg
Real ID
Text
Word ID
Wordbase
Wordform
Measure Names
City Record
COG Avg
Cog1
Cog2
Cog3
Content
Decade
DOB
Education
Fam County
Fam South
Fam State

Pages

Filters

Exclusions (DOB, T..)

Marks

Automatic

Color Size Label

Detail Tooltip

Task

Columns

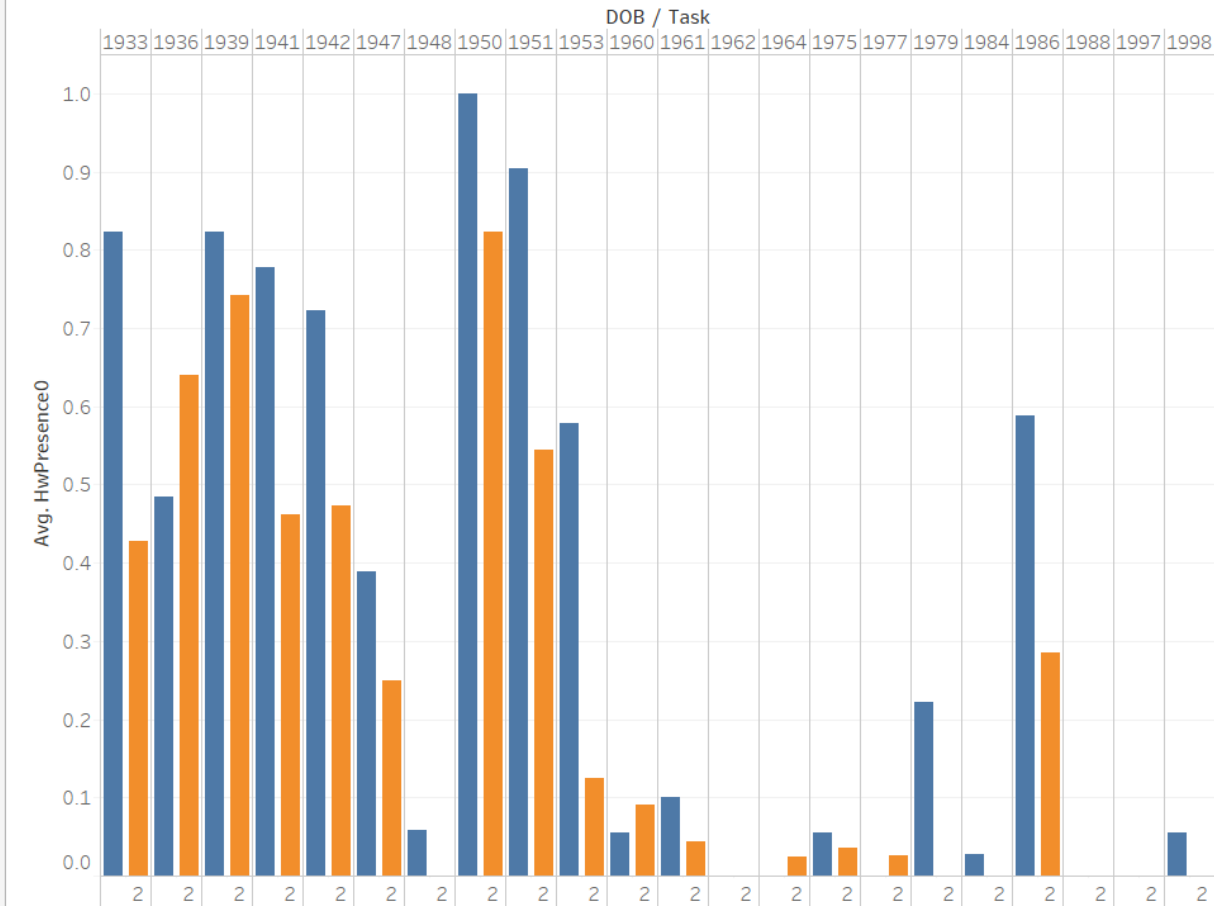
DOB

Task

Rows

AVG(HwPresence0)

Sheet 1



Show Me



For horizontal bars try

0 or more Dimensions

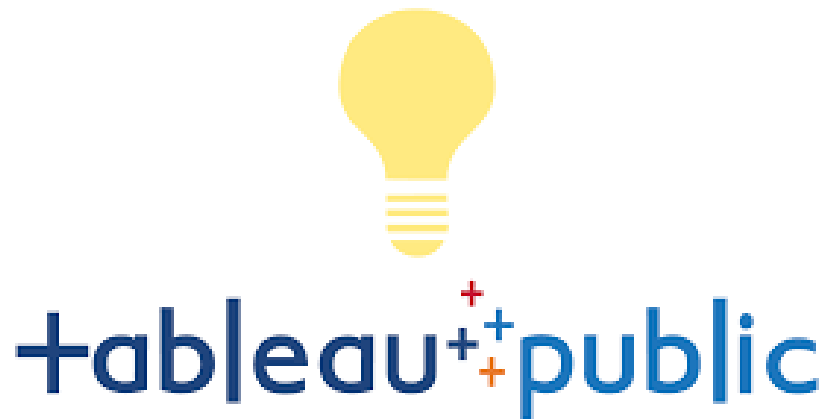
1 or more Measures

Data Source

Sheet 1

44 marks 1 row by 44 columns SUM of AVG(HwPresence0): 12.663

Type here to search



- Offers many different beautiful chart options:
 - Scatterplot
 - Heatmap
 - Barplot
 - Box-and-whisker plot
 - Histogram
 - Data dashboard(s)

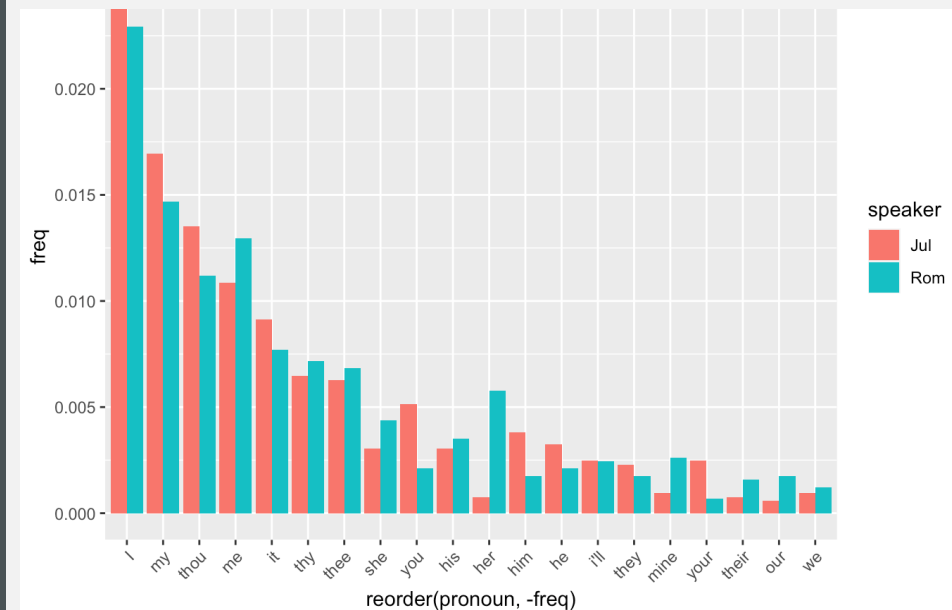
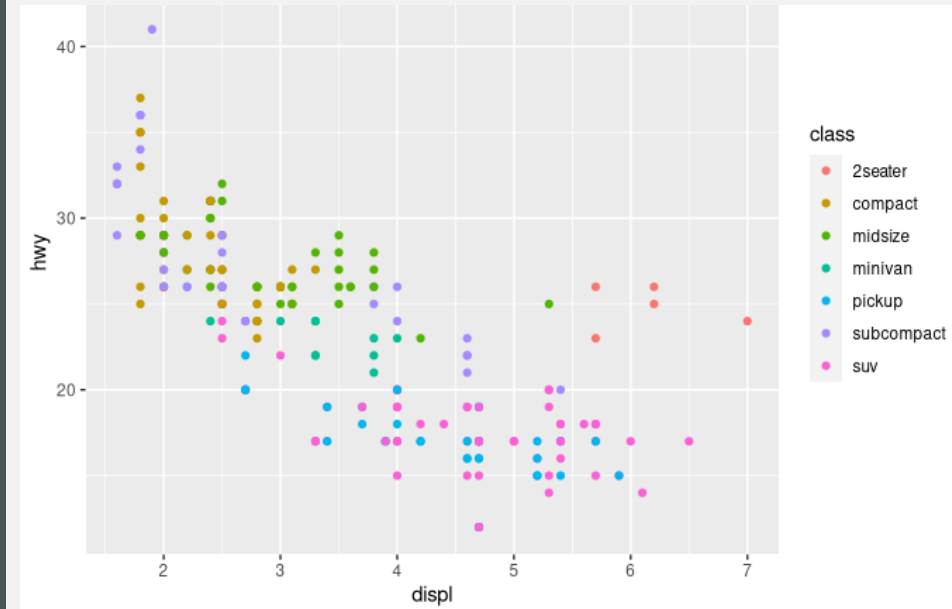
- Examples:

<https://public.tableau.com/en-us/gallery/shark-attacks-us>

<https://public.tableau.com/en-us/gallery/must-read-books-black-authors>



- General graphics packages:
 - **Base/standard graphics:** provided by the graphics package in base R
 - **Trellis/lattice graphics:** provided by the lattice package
 - **ggplot2** based on Wilkinson's *Grammar of Graphics* and available from CRAN



- Interactive packages:
 - **Leaflet**- interactive maps
 - **Plotly**- useful for interactive plots and other graphics including 3D charts
 - **SunburstR**- sequence charts
 - **Dygraphs**- useful for time series data
 - **Iplots**- a Java-based dynamic graphics system for linked plots
 - **Shiny** - provides the opportunity to create dynamic, interactive visualizations, as shown in this [Gazetteer of Southern Vowels](#), by J. Stanley



PYTHON

- **Matplotlib** – most widely used library for plotting in Python, designed to resemble MATLAB, and works well in conjunction with other libraries like pandas
- **Seaborn**- has more color and default options and builds into matplotlib
- **Ggplot**- integrates well with pandas, plot components are implemented in layers create a complete plot, ported from R
- **Bokeh**- creates interactive, web-ready plots that are easily output as JSON objects, html, or interactive web apps
- **Plotly**- great for interactive plots, and makes contour plots, dendograms, and 3D charts
- **Geoplotlib**- toolbox for geographical data and mapping, including dot density maps, heatmaps, and choropleths



RESOURCES

- Other web-based resources:
 - [DataBasic.io](#)
 - [RAWGraphs](#)
 - [InfoGram](#)
 - [Datacopla](#)
 - [440 Ways to Visualize Text Data](#) the Text Visualization Browser, ISOVIS group
 - [Machine Learning Models and Visualizations](#), ISOVIS group
- ArcGIS and QGIS are useful tools for creating maps.



RECOMMENDED READING

- *Visualizing Data and Elements of Graphing Data* by William Cleveland
- *The Visual Display of Quantitative Information, Visual Explanations, Envisioning Information, Beautiful Evidence* by Edward Tufte
- *Information is Beautiful* by David McCandless
- *The Visualization Toolkit* by Schroeder et al.

RECOMMENDED READING

- *Design Methods* by Amy Ko
- Hadley Wickam's *A Layered Grammar of Graphics*
- *Bad Data Guide* by Quartz Data Team
- Heer et al. *A tour through the visualization zoo*
- *Cartographies of Time: A history of the Timeline* by Rosenberg and Grafton
- *Visual Thinking for Design* by Colin Ware
- For more on GIS resources, check out this article by Emily McGinn and Meagan Duever : *We mapped it so you don't have to: comparing online data mapping platforms*

RECOMMENDED WEBSITES

- Excellent overview of selecting different types of graphs for different types of data!
- Information is beautiful by David McCandless
- Edward Tufte's website
- Excel chart cookbook from Duke Libraries
- <https://datavizcatalogue.com>



RECOMMENDED WEBSITES



[Visualizing text data with ggplot2](#)



[Ggplot2 cheat sheet](#)



[Mastering Shiny by Hadley Wickham](#)



[Tidyverse Blog](#)

ADDITIONAL RESOURCES

- Linked-in learning has many tutorials for data visualization in excel and R.
- Tableau also has freely available tutorials.
- Tutorials from the Digilab here at UGA: excel, R, and Tableau!
- R Ladies Athens!!
- Courses at UGA
- DigiLab Data Office Hours

DATA OFFICE HOURS



CONSULTATIONS FOR DATA CLEANING, STRUCTURING, AND VISUALIZING

Whether just starting your work, or trying to make sense of your research, schedule an appointment for our Data Office Hours and bring your data (text, archival information, numerical data, etc.) for advice and guidance on your project. Expertise in corpus linguistics, Excel, and R, among other tools for data structuring and visualization.

TUESDAYS • 4:00-5:00
WEDNESDAYS • 2:00-3:00

To schedule an appointment visit:
DIGI.UGA.EDU/RESOURCES

WILLSON
CENTER **DIGILAB**



THANKS FOR LISTENING!

KATHERINE.KUIPER25@UGA.EDU
KEIKO.BRIDWELL@UGA.EDU

WORKS CITED

- Cash, Ivan. <https://i.imgur.com/DNxvj.png>.
- Chang, Agnes. 2020. Introduction to Data Visualization: Course Syllabus. Columbiaviz.github.io
- Cieslak, Jack. 13 Powerful ways to visualize your data. Sisense.com
- Durcevel, Sandra. Designing Charts and Graphs: How to Choose the Right Data visualization Types. May 2nd, 2019. datapine.
- <https://www.smithsonianmag.com/history/surprising-history-infographic-180959563/>
- Durcevel, Sandra. Designing Charts and Graphs: How to Choose the Right Data visualization Types. May 2nd, 2019. datapine.
- ETA+. 2020. Unsplash.
- Friendly, Michael. Historical Milestones. <http://www.datavis.ca/>
- Friendly, Michael. 2008, "The Golden Age of Statistical Graphics," *Statistical Science* 8(4), 502—535
- <https://ggplot2.tidyverse.org>
- Gupta, Sujata. 2020. Florence Nightingale Understood the Power of Visualizing Science. *Science News*. <https://www.sciencenews.org/article/florence-nightingale-birthday-power-visualizing-science>
- Harrell, J.A. and V.M. Brown, 1992a, "The world's oldest surviving geological map – the 1150 BC Turin papyrus from Egypt", *Journal of Geology* 100 (1992), pp.3–18. Harrell, J.A. and V.M. Brown, 1992b, "The oldest surviving topographical map from ancient Egypt (Turin Papyri 1879, 1899 and 1969)", *Journal of the American Research Center in Egypt* 29 (1992), pp.81–105. Klemm,
- Hill, Asha. 2018. 9 Useful R Data Visualization Packages for Any Discipline. Mode. <https://mode.com/blog/r-data-visualization-packages/>
- Kiss, Zoltan. 2018. Data and Data Analysis in Linguistics. <http://seas3.elte.hu/lingtheo/notes/gkiss-data-in-lcs.html>.
- Klemm, Rosemarie; Klemm, Dietrich (2013). *Gold and Gold Mining in Ancient Egypt and Nubia*. Heidelberg: Springer. pp. 132–136, 144
- Kretzschmar et al. 2012. Digital Archive of Southern Speech.
- Kretzschmar, W. and I. Juuso. 2014. Cellular Automata for Modeling Language Change. International Conference on Cellular Automata.
- Lam, Longhow. 2017. Interactive sunburst graphs in Power BI in 5 minutes!. <https://www.r-bloggers.com/2017/07/interactive-sunburst-graphs-in-power-bi-in-5-minutes/>
- Open Learn. 2020. Session 3: Discrete and Continuous Data. Openlearn Edu: Uk. <https://www.open.edu/openlearn/ocw/mod/oucontent/view.php?id=85587§ion=1>
- <https://vocal.media/futurism/most-powerful-star-wars-jedi>
- Simmons, Ryan. 2016. ENC:2135 Research, Genre & Context. <https://dfb1585.wixsite.com/enc2135/single-post/2016/11/10/Blog-Prompt-1>
- Spiske, Markus. 2020. Unsplash
- Stephanie. 2020. Statistics How To: Statistics for the rest of us!. Types of Variable.
- Stanley, J. W. Kretzschmar, M. Renwick, M. Olsen, and R. Olsen. 2017. *Gazetteer of Southern Vowels*. Linguistic Atlas Project, University of Georgia.
- Tableau. 2020. Data visualization beginner's guide: a definition, examples, and learning resources. <https://www.tableau.com/learn/articles/data-visualization>
- Tufte, Edward. 1983, *The Visual Display of Quantitative Information*.
- Tufte, Edward. 1990. *Envisioning Information*.
- Tufte, Edward. 2006. *Beautiful Evidence*.
- "Turin Papyrus Map from Ancient Egypt", *Research on the Archaeological Geology of Ancient Egypt*, Accessed Oct 26, 2020, http://www.eeescience.utoledo.edu/Faculty/Harrell/Egypt/Turin%20Papyrus/Harrell_Papyrus_Map_text.htm.
- UC Denver Libraries. Research Guides: auraria Library. <https://guides.auraria.edu/c.php?g=528873&p=3964258>.
- Wickam, Hadley. 2010. A layered grammar of graphics. *Journal of Computational and Graphical Statistics*.
- Wiediger, Christian. 2020. Unsplash
- Wilson, Mark. 2017. The real difference btewen Google and Apple. *Fast Company*. <https://www.fastcompany.com/3068474/the-real-difference-between-google-and-apple>