# Data Visualization Workshop

David Sasson / Wei-Hung Weng

HST.953 Collaborative Data Science in Medicine Sep 14, 2018



# Agenda

- Quick introduction
- Data Visualization with R and ggplot2
  - Powerful visualization tool in R
- Visualization for clinical applications
  - Applied visualization with simple R commands

# Why Visualization?

- A method of encoding quantitative, relational, or spatial information into images
- Taps into the visual system an enormously powerful pattern-finding device –
   which can reveal structure in data in a compelling and accessible way

David Sasson

#### Goal of Visualization

 The greatest value of a picture is when it forces us to notice what we never expected to see.

John Tukey (1977)

The purpose of visualization is insight, not pictures.

Ben Shneiderman (1999)

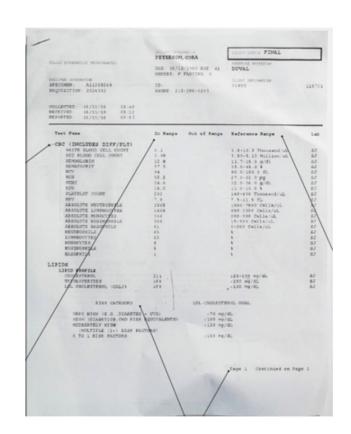
- Understanding and exploring trends and patterns inside data
- Summarizing statistics
  - Instead of reading thousands raw data points
- Telling a story!

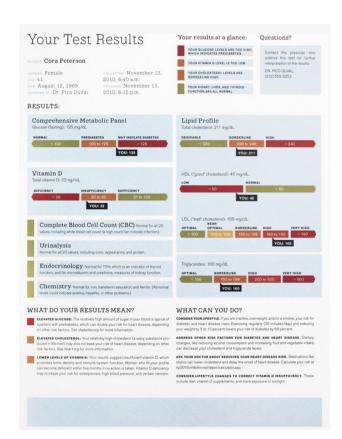
#### Medical Data?

Why challenging (for analysis as well as visualization)?

- Volume
- Missing data
- Trusting source of data/resolving conflicting data
- Time series
- Change/acceleration vs. absolute (whether in spending or in disease progression)
- Bias

#### Do This

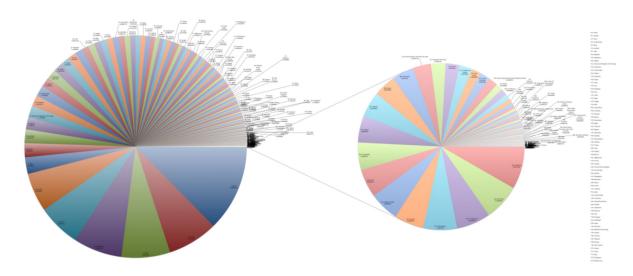




## Don't Do This!

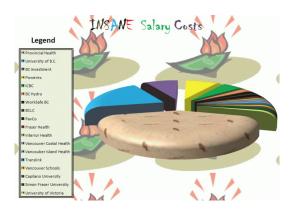
http://viz.wtf

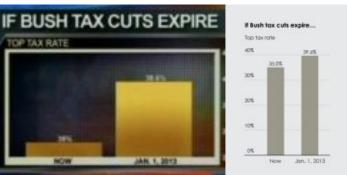
http://eagerpies.com/close-the-bars-down/

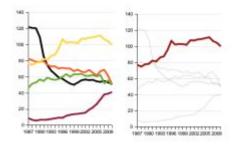


## Some Principles

- Lets the data speak for itself
- The addition of extra fluff (shadows, 3D, extravagant colors) eclipses what the graph is actually showing
- Faithful to the data, and doesn't misrepresent it by modifying axes or colors the wrong way
- Data visualization is as much of an art as it is a science
- Minimalism







#### 10 Commandments of Data Visualization

DO

Use the full axis

Simplify less important information

Be creative with legends and labels

Utilize a hierarchy

Ask others for opinions

**DON'T** 

Use 3D effects

Use more than six colors

Change visual style

Make people do visual math

Overload the chart

David Sasson

## Some Tips...

- Show the data
- Induce the viewer to think about the substance, rather than about methodology, graphic design, [or] the technology of graphic productions...
- Avoid distorting what the data have to say
- Present many numbers in a small space
- Make large data sets coherent
- Encourage the eye to compare different pieces of data
- Reveal the data at several levels of detail
- Serve a reasonably clear purpose
- Be closely integrated with the statistical and verbal descriptions

Edward Tufte, The Display of Quantitative Information

## More Tips...

- Written things proceed from left to right (in English)
- Things proceed from top to bottom
- Center things are more important than periphery things
- Foreground things are more important than background things
- Thick things are more important than thin things
- Areas of activity contain the most important information
- Things with the same shape, size, color, or location are related
- Things stand out if they contrast with surroundings in terms of line thickness, type face, or color

T. Huckin and L. Olsen, English for Science and Technology

## **Further Readings**

- Harvard CS171 Visualization
  - o <a href="http://www.cs171.org">http://www.cs171.org</a>
- GaTech CS 7450 Information Visualization
  - https://www.cc.gatech.edu/~stasko/7450/
- Edward Tufte
  - https://www.edwardtufte.com/tufte/
- David McCandless
  - https://informationisbeautiful.net/
- Toolkits
  - o D3JS
  - R Shiny
  - Tableau (especially if you use Google BigQuery)
  - http://selection.datavisualization.ch/

#### Connect Microsoft Excel Text file JSON file PDF file Spatial file Statistical file More... Tableau Server MySQL Oracle Amazon Redshift Google BigQuery More... Sample - Superstore World Indicators

Tableau

Open

Sample Workbooks

Superstore

.

Open a Workbook

Training

Discove

Getting St

Connectin Visual Ana

Understan

More train

% Sharing

Learn mor

Resource

Get Tablea

Blog - Wor Services in MATLAB

Register fo

Forums

European Foot Club Rankings

More Samples

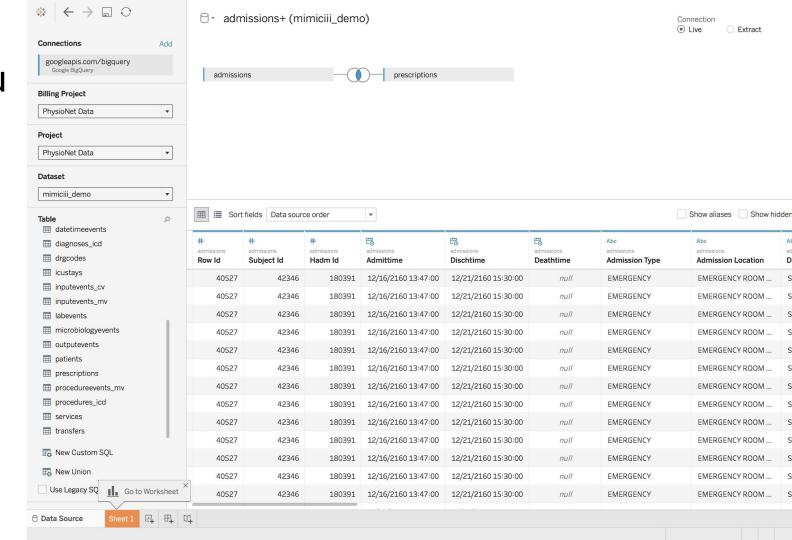


World Indicators



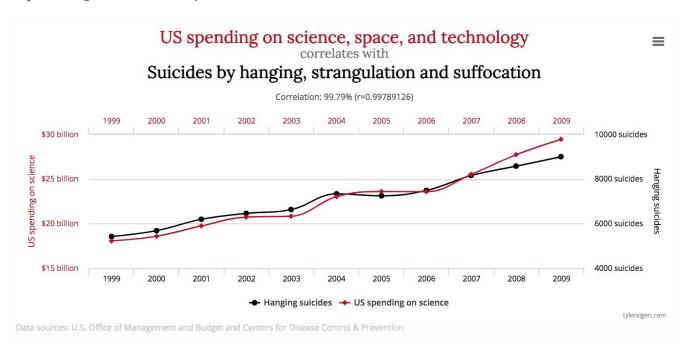
Regional

#### Tableau



## **Next Step**

http://tylervigen.com/spurious-correlations



## Tutorial - Introduction to ggplot2

- You need R and RStudio
- https://github.com/ckbjimmy/hst953\_viz
  - Clone or download
- https://github.com/dsasson48/dataviz

# Tutorial - Visualization for Clinical Applications

- You need R and RStudio
- https://github.com/ckbjimmy/hst953\_viz
  - Clone or download
- What's inside the tutorial? Plotting Function in R
  - Histogram
  - Density estimation
  - Scatter plot
  - Boxplot
  - Interaction plot
  - Supervised visualization
  - Model validation
  - Summarization
  - Other issues...