Visualizing the World

An Introduction to Visualization
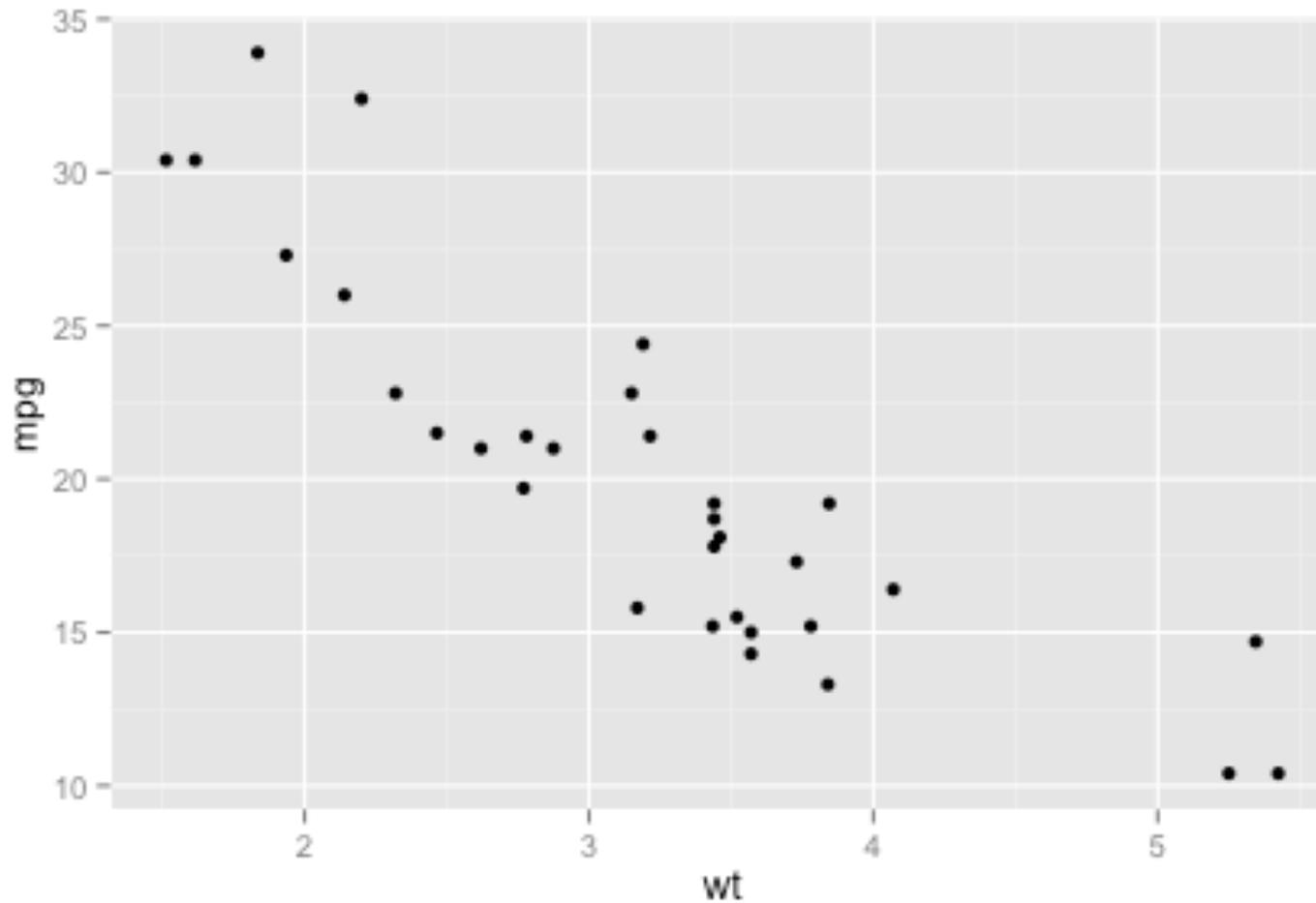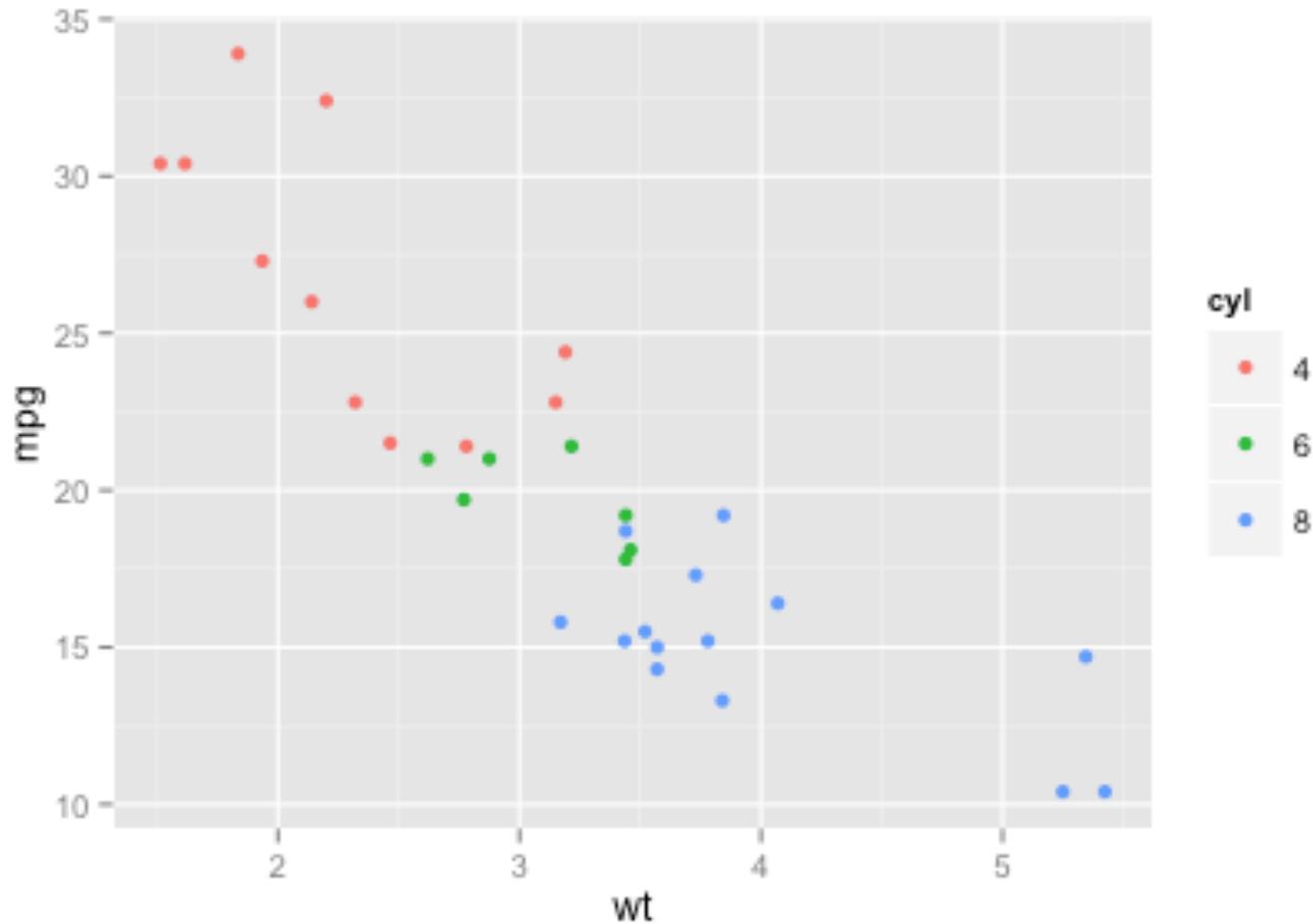
Why Visualization?

- "The picture-examining eye is the best finder we have of the wholly unanticipated"

  -John Tukey

- Visualizing data allows us to discern relationships, structures, distributions, outliers, patterns, behaviors, dependencies, and outcomes

- Useful for initial data exploration, for interpreting your model, and for communicating your results
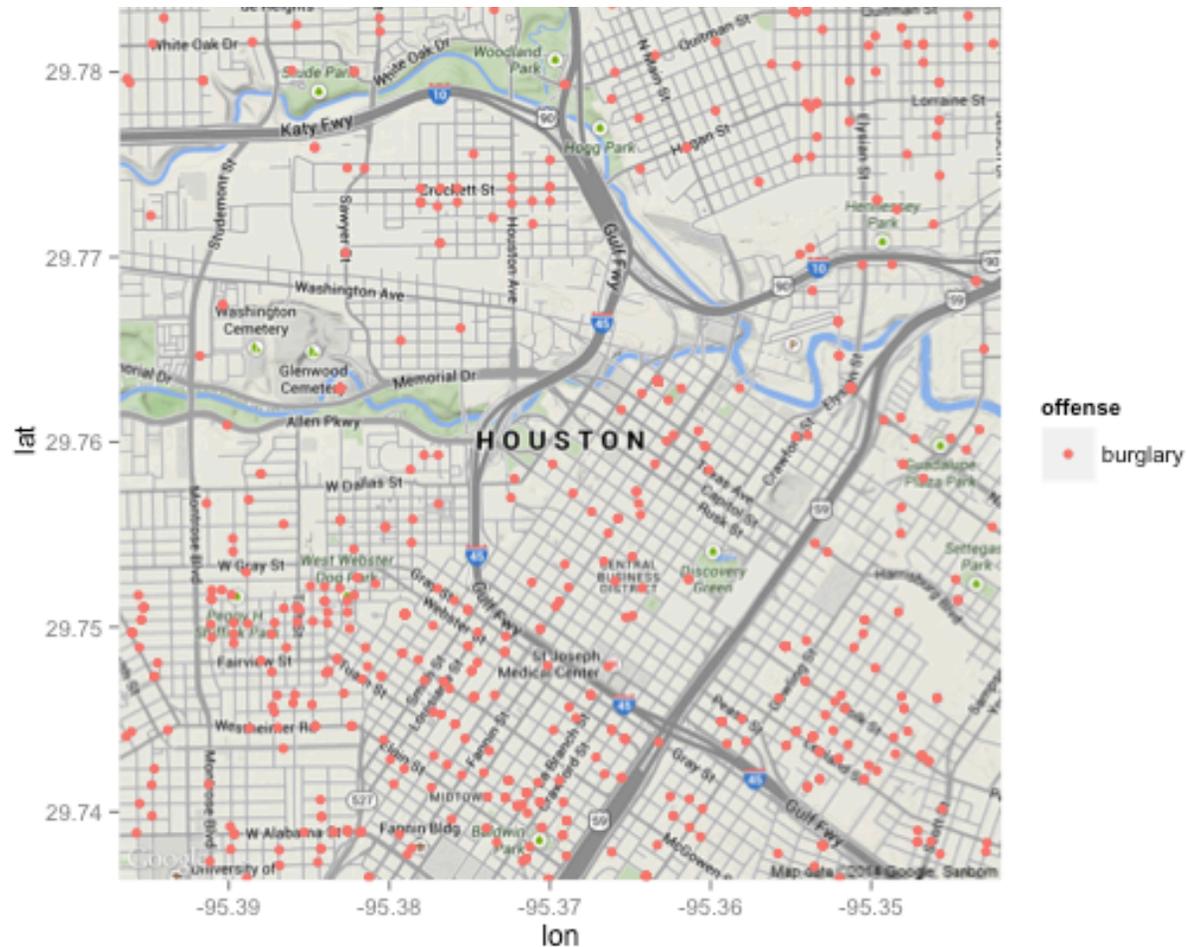
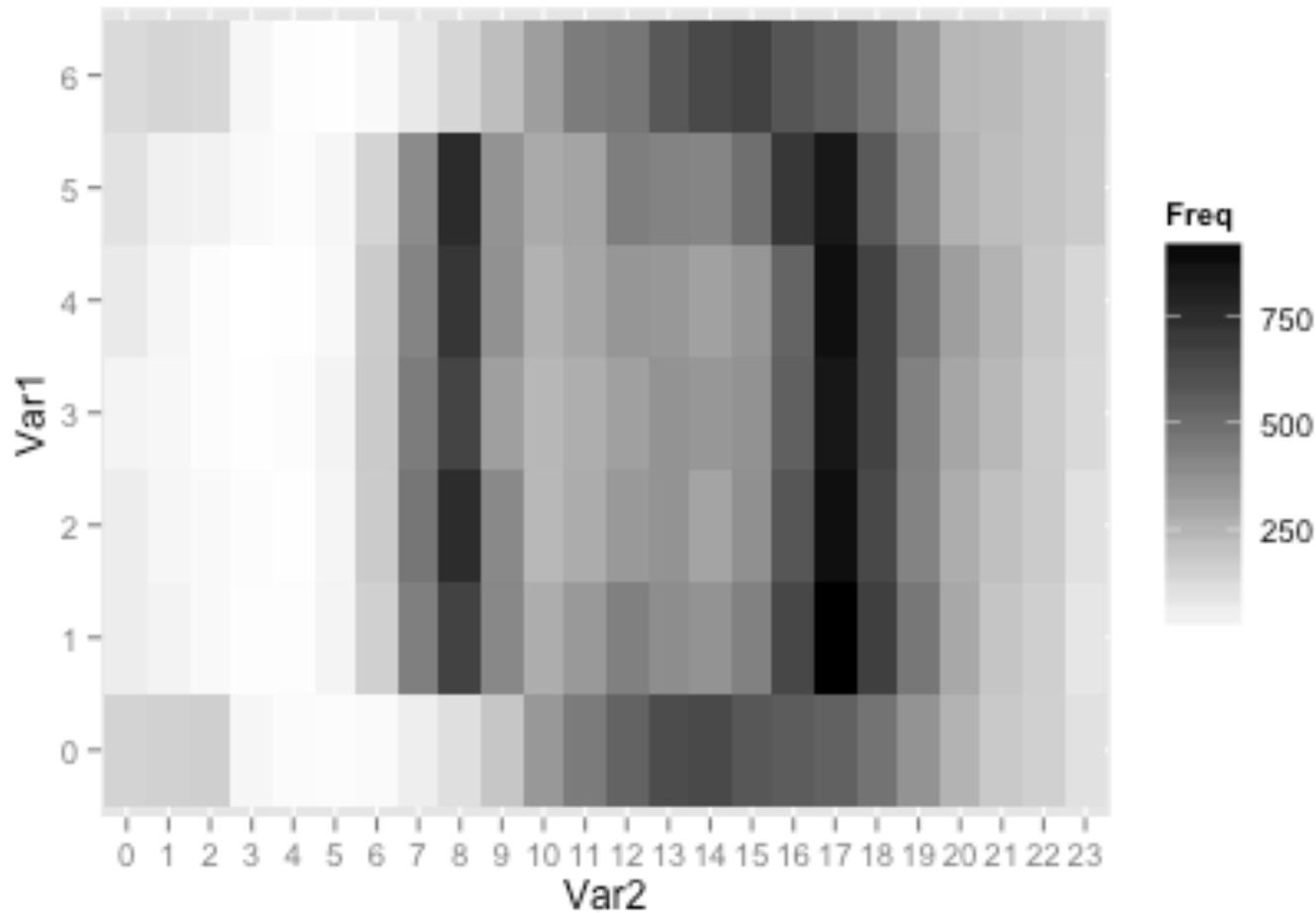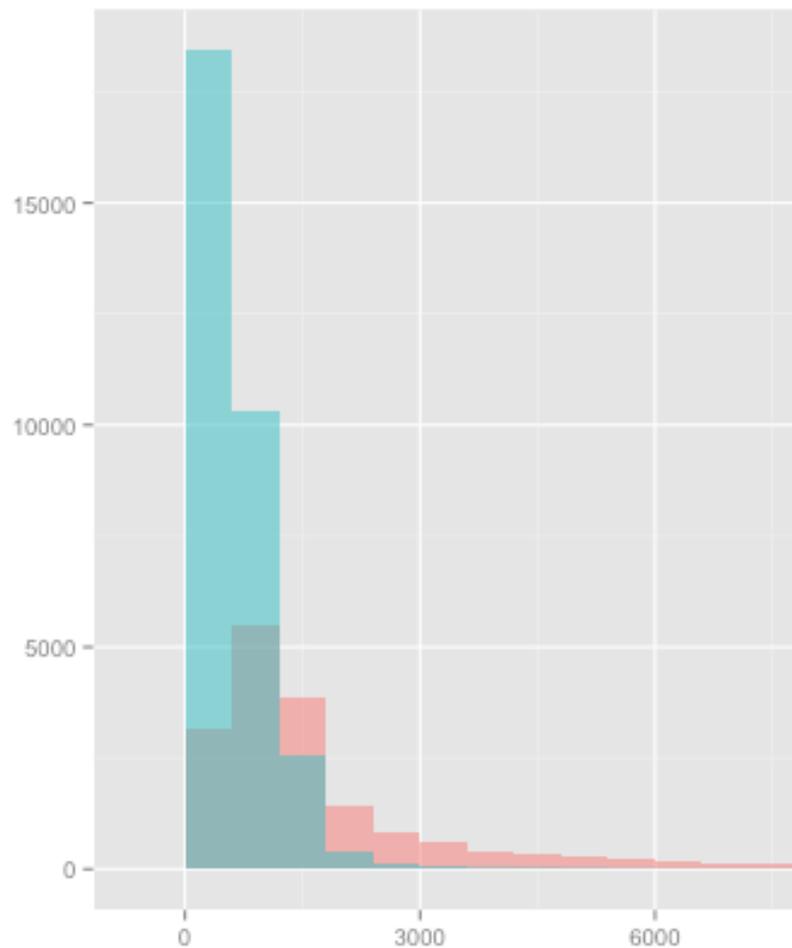# Initial Exploration Shows a Relationship

Explore Further: Color by Factor

Make a Model. Plot the Regression Line.

Add Geographical Data to a Map

Show Relationships in a Heatmap

Make Histograms. Explore Categories.

Color a Map According to Data
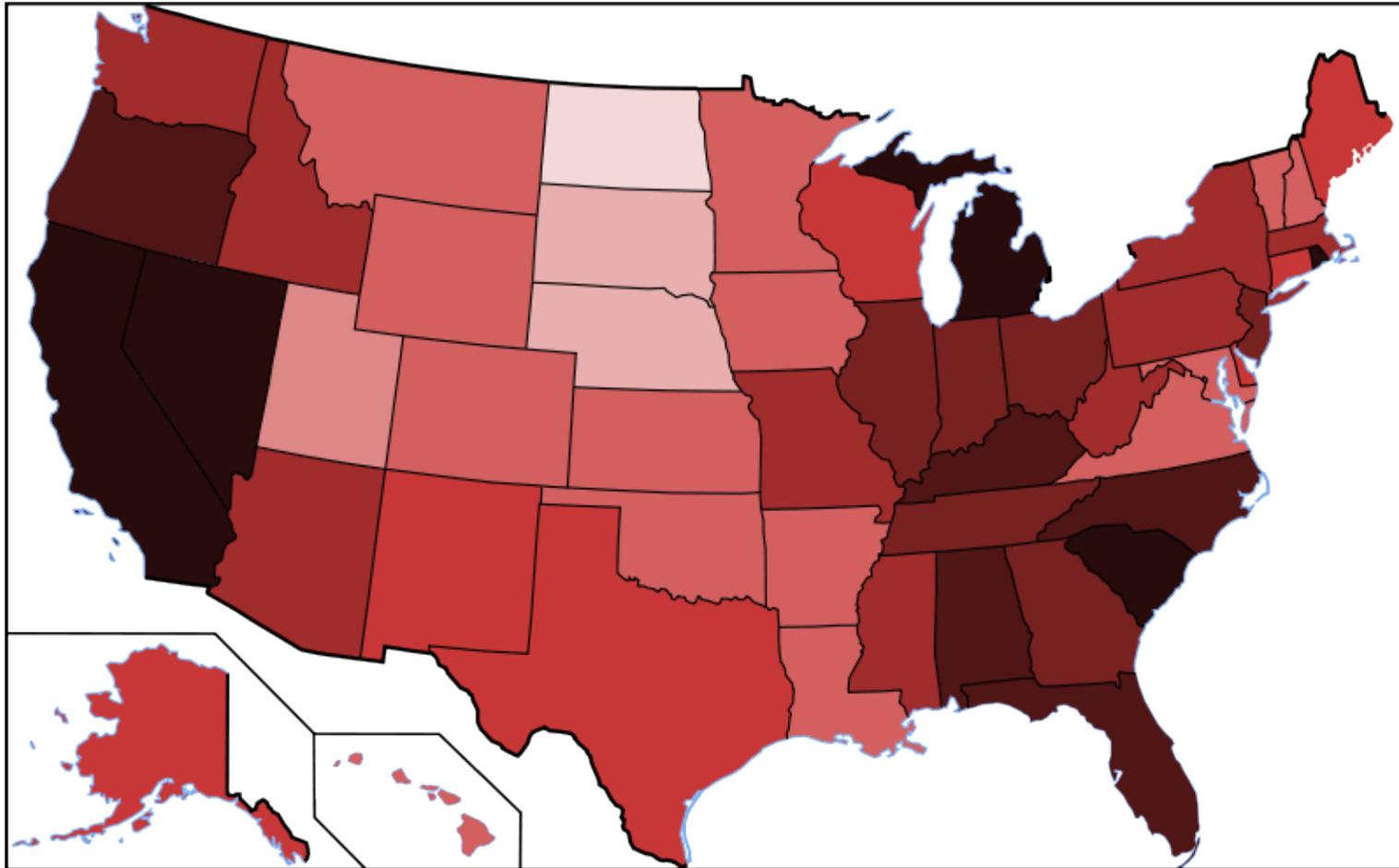
The Power of Visualizations

- This week, we will create all of these visualizations

- We will see how visualizations can be used to
  - Better understand data
  - Communicate information to the public
  - Show the results of analytical models

- In the next video, we will discuss the World Health Organization (WHO), and how they use visualizations

The World Health Organization

"WHO is the authority for health within the United Nations system. It is responsible for providing leadership on global health matters, shaping the health research agenda, setting norms and standards, articulating evidence-based policy options, providing technical support to countries and monitoring and assessing health trends."

The World Health Report

- WHO communicates information about global health in order to inform citizens, donors, policymakers and organizations across the world

- Their primary publication is "World Health Report"

- Each issue focuses on a specific aspect of global health, and includes statistics and experts' assessments

Online Data Repository

- WHO also maintains an open, online repository of global health data

- WHO provides some data visualizations, which helps them communicate more effectively with the public

World Energy
Consumption,
2001-2003
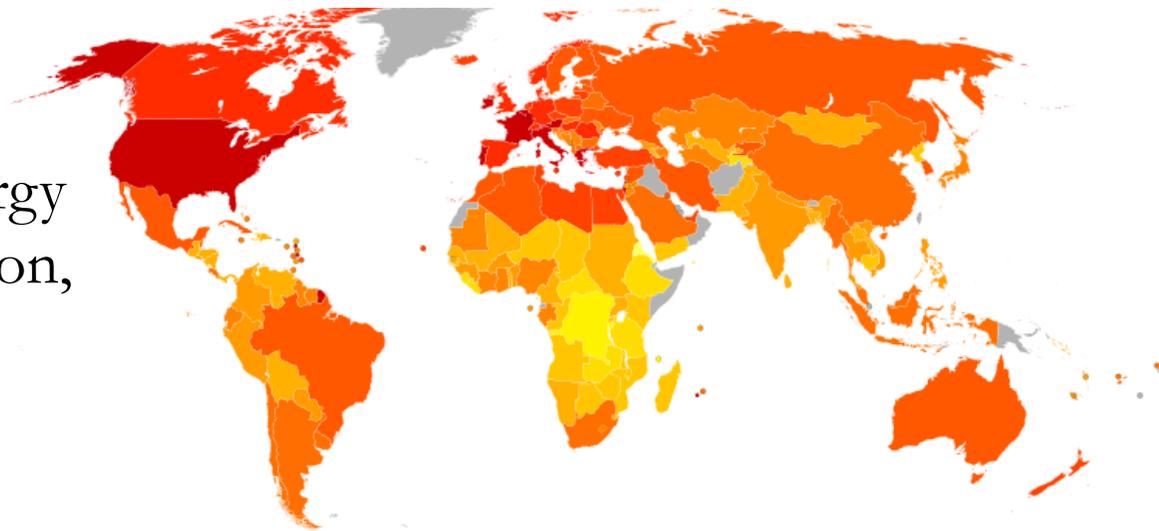
What is a Data Visualization?

- A mapping of data properties to visual properties

- Data properties are usually numerical or categorical

- Visual properties can be (x,y) coordinates, colors, sizes, shapes, heights, . . .

Anscombe's Quartet

X1	Y1	X2	Y2	X3	Y3	X4	Y4
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

- Mean of X: 9.0
- Variance of X: 11.0
- Mean of Y: 7.50
- Variance of Y: 4.12
- Correlation between X and Y: 0.816
- Regression Equation:
- $Y = 3.00 + 0.500X$

Anscombe's Quartet
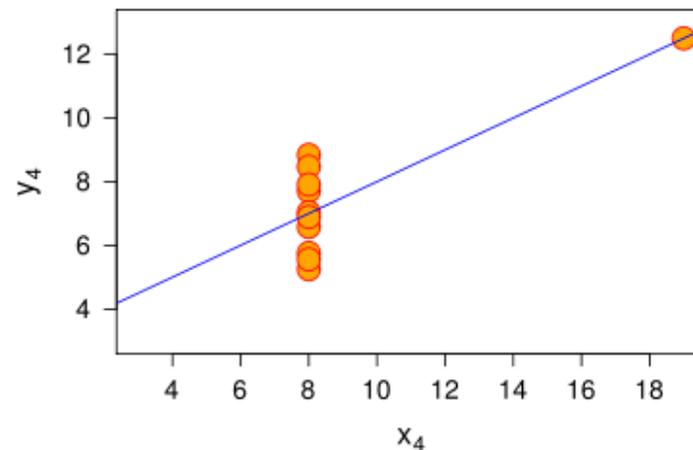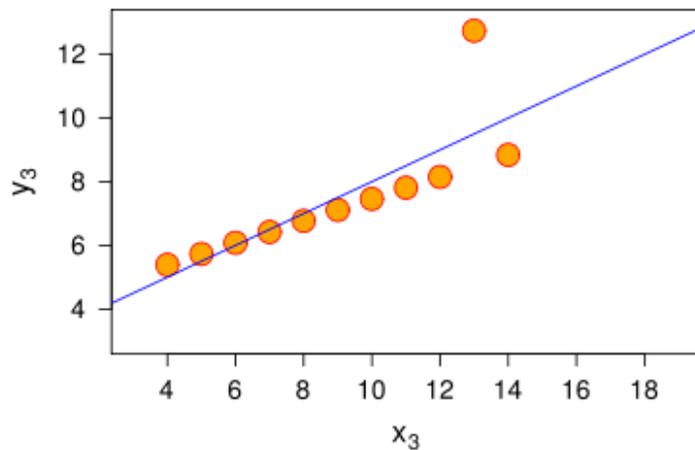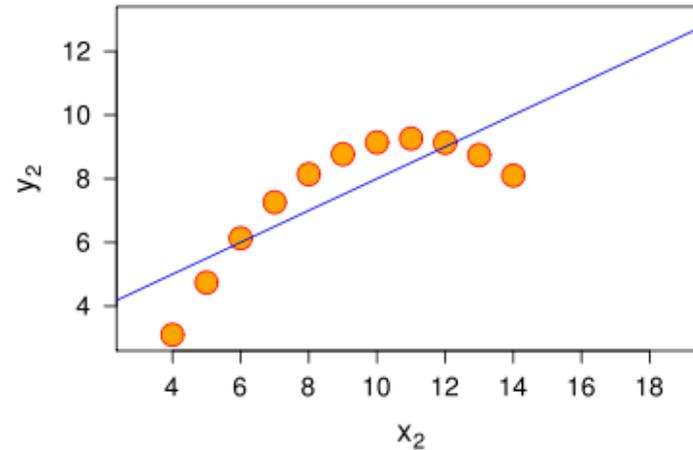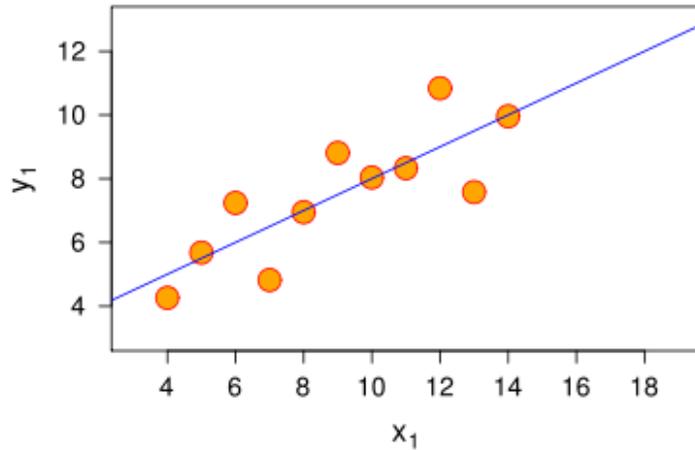
ggplot

- "ggplot2 is a plotting system for R, based on the grammar of graphics, which tries to take the good parts of base and lattice graphics and none of the bad parts. It takes care of many of the fiddly details that make plotting a hassle (like drawing legends) as well as providing a powerful model of graphics that makes it easy to produce complex multi-layered graphics."

-Hadley Wickham, creator, www.ggplot2.org

Graphics in Base R vs ggplot

- In base R, each mapping of data properties to visual properties is its own special case

  - Graphics composed of simple elements like points, lines
  - Difficult to add elements to existing plots

- In ggplot, the mapping of data properties to visual properties is done by adding layers to the plot

Grammar of Graphics

- ggplot graphics consist of at least 3 elements:

  1. **Data**, in a data frame

  2. **Aesthetic mapping** describing how variables in the data frame are mapped to graphical attributes

     ➢ Color, shape, scale, x-y axes, subsets,…

  3. **Geometric objects** determine how values are rendered graphically

     ➢ Points, lines, boxplots, bars, polygons,…

The Analytics Edge

- WHO's online data repository of global health information is used by citizens, policymakers, and organizations across the world

- Visualizing the data facilitates the understanding and communication of global health trends at a glance

- ggplot in R lets you visualize for exploration, modeling, and sharing results