

# Welcome

Statistics 4868/6610 Data Visualization

Prof. Eric A. Suess

January 3, 2018

## Introduction

Why am I interested in Data Visualization?

Why are you interested in Data Visualization?

**Data Viz** is very cool these days!

- data
- summarize it
- make a picture
- tell a story

## Former students

- Yimeng Wang Women at the Wheel
- Jeff Hemsley [somalab.net](http://somalab.net)
- Aida Yazdanparast and Tony Tran Oil Spill on Birds
- Prashanth Chandrappa Master's Project - Sentiment Analysis of Tweets Using Hash Tags
- Joseph Rickert founder of BARUG
- Gabriela de Queiroz founder of R-Ladies SF

## Former students

RStudio - Joe Rickert - revoJoe - BARUG

Sharethrough - Gabriela de Querioz - RLadiesSF - RLadies

## Cabinet Exit Memos

President Obama asked each member of his Cabinet to write an Exit Memo on the progress made.

President Obama's Exit Memo

He has made effective use of data visualization. IMO

## This course

In this course we are going to examine Data Visualization examples, methods, and software.

You will have the opportunity to work with data to create visualizations.

## The book

Visualize This: The FlowingData Guide to Design, Visualization, and Statistics by Nathan Yau

## The book's website

FlowingData

See Tutorials and Projects

## Websites

youtube

- Hans Rosling BBC
- Hans Rosling: The best stats you've ever seen
- The beauty of data visualization - David McCandless
- tableau videos

## Software

What **software** is used for Data Visualization?

- R and R Studio
- R Studio shiny
- ggobi
- Tableau
- Plotly
- Python and matplotlib and Seaborn and Bokeh
- D3

## Software

- MS Excel
- JMP
- SAS
- Minitab
- SPSS

and many many more ...

blog post Top Data Visualization Software Products

## History

- Exploratory Data Analysis (EDA)
- Exploratory Data Analysis by John Tukey

## Other books

- The Visual Display of Quantitative Information by Edward R. Tufte
- The Grammar of Graphics by Leland Wilkinson
- ggplot2 by Hadley Wickham
- Graphics of Large Datasets, Visualizing a Million by Unwin, Theus, Hofmann

## The cloud

Do you know what **The Cloud** is?

Why is it so important for data visualization?

## Simple R example

```
summary(cars)
```

speed		dist	
Min.	: 4.0	Min.	: 2.00
1st Qu.	:12.0	1st Qu.	: 26.00
Median	:15.0	Median	: 36.00
Mean	:15.4	Mean	: 42.98
3rd Qu.	:19.0	3rd Qu.	: 56.00
Max.	:25.0	Max.	:120.00

## Simple R example

```
plot(cars)
```

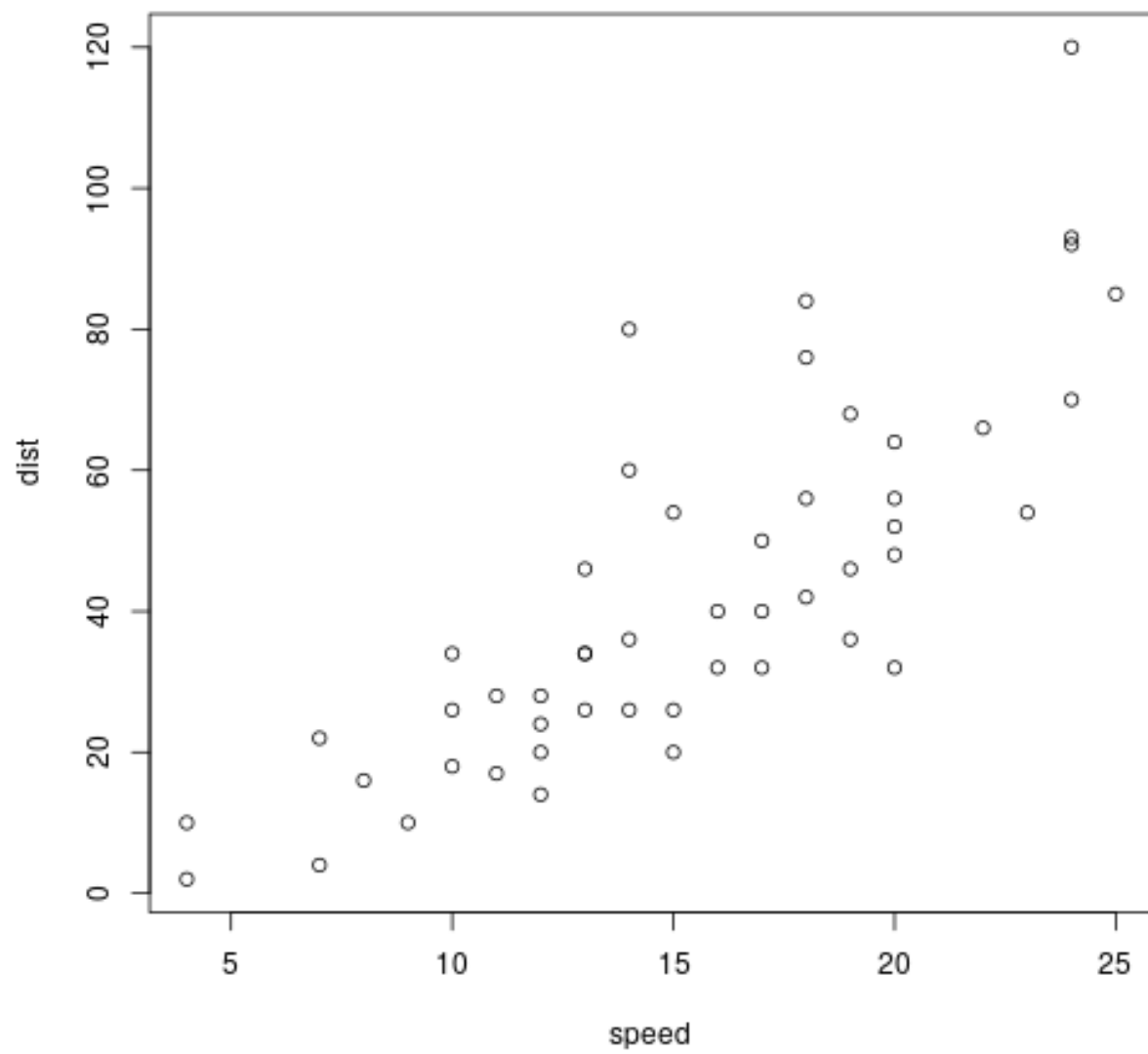


Figure 1: plot of chunk unnamed-chunk-2