

# DataTypes

Statistics 4868/6610 Data Visualization

Prof. Eric A. Suess

1/13/2016

## Introduction

Topics today.

- Design
- data formats - .xlsx, .txt, .json, .xml
- DataSF
- Are you a Hacker, Scripter, or Application Users, or all 3?
- Where did tableau come from?
- What is D3?
- kaggle

## Design

**Question:** So what is the underlying design of the websites we have looked at?

- [yahoo finance](#)
- [Weather Underground](#)

## Design

**Answer:** High level overview goes into more and more detail until you can change the views yourself.

## DataSF

Back to the [DataSF](#) website.

What kind of data files can you download?

Which areas of SF have smart meters? See the Parking meters dataset.

Another website with SF data [SFpark](#)

## Data and the web

If you are going to work with data and the web, you are likely to run into data in JSON and XML formats.

- Javascript object notation (JSON)
- Extensible Markup Language (XML)
- XML is commonly used with APIs
- Application Programming Interface (API)

So if you going into writing apps, ...

## Enterprise Data Analysis and Visualization

Recently there was an interview study conducted by some faculty at Stanford and UC Berkely to examine the differences in companies between different kinds of data analysis.

**Question:** Which of the following **three archetypes** are you?

- Hackers
- Scripters
- Application Users

## Enterprise Data Analysis and Visualization

Read the Section 6 Future Trends.

[Enterprise Data Analysis and Visualization: An Interview Study](#) by Kandel, Paepcke, Hellerstein, Heer

Related

An interesting related TEDx talk [Data hacking - data science for entrepreneurs | Kevin Novak | Uber](#)

PBS kids [cyberchase](#) in the show they are [Using Data](#)

Is the Data Hacker now the Data Scientist?

## Where did tableau come from?

Short answer, Stanford

[Polaris: A System for Query, Analysis and Visualization of Multidimensional Relational Databases](#) by Stolte, Tang, Hanrahan

## Where did D3: Data Drive Documents come from?

Seems to be Stanford and/or the University of Washington [idl](#)

[D3](#)

[github/mbostock](#)

[D3.js gallery](#)

## kaggle

Data competitions.

[Bike Sharing Demand](#)

Download the data. Visualize it!

## github

So what is github? You might have heard of it. It is a very useful cloud based platform for sharing code.

[github](#)

[github/esuess](#)

## Useful tools

Some useful tools mentioned in the book. Shared on github.

- [OpenRefine 2014 survey results](#)
- [Mr. Data Coverter](#)
- [Mr. People](#)

## podcasts RSS and XML

An audio podcast that discusses visualization.

[datastori.es](#)

An audio podcast about R

[The R-Podcast](#)

Local online TV with a nice show about learning to code.

[twit.tv Coding 101](#)

Microsoft TV

[Channel 9](#)

## How listen to music

When I listen to music or podcasts I often listen using

[Sonic Visualiser](#)

The aim of Sonic Visualiser is to be the first program you reach for when want to study a musical recording rather than simply listen to it.

# Wrapping Up

Find data

Manage data

Visualization process

Okay, you have your data. Now it's time to get visual!

# Slide With R Code

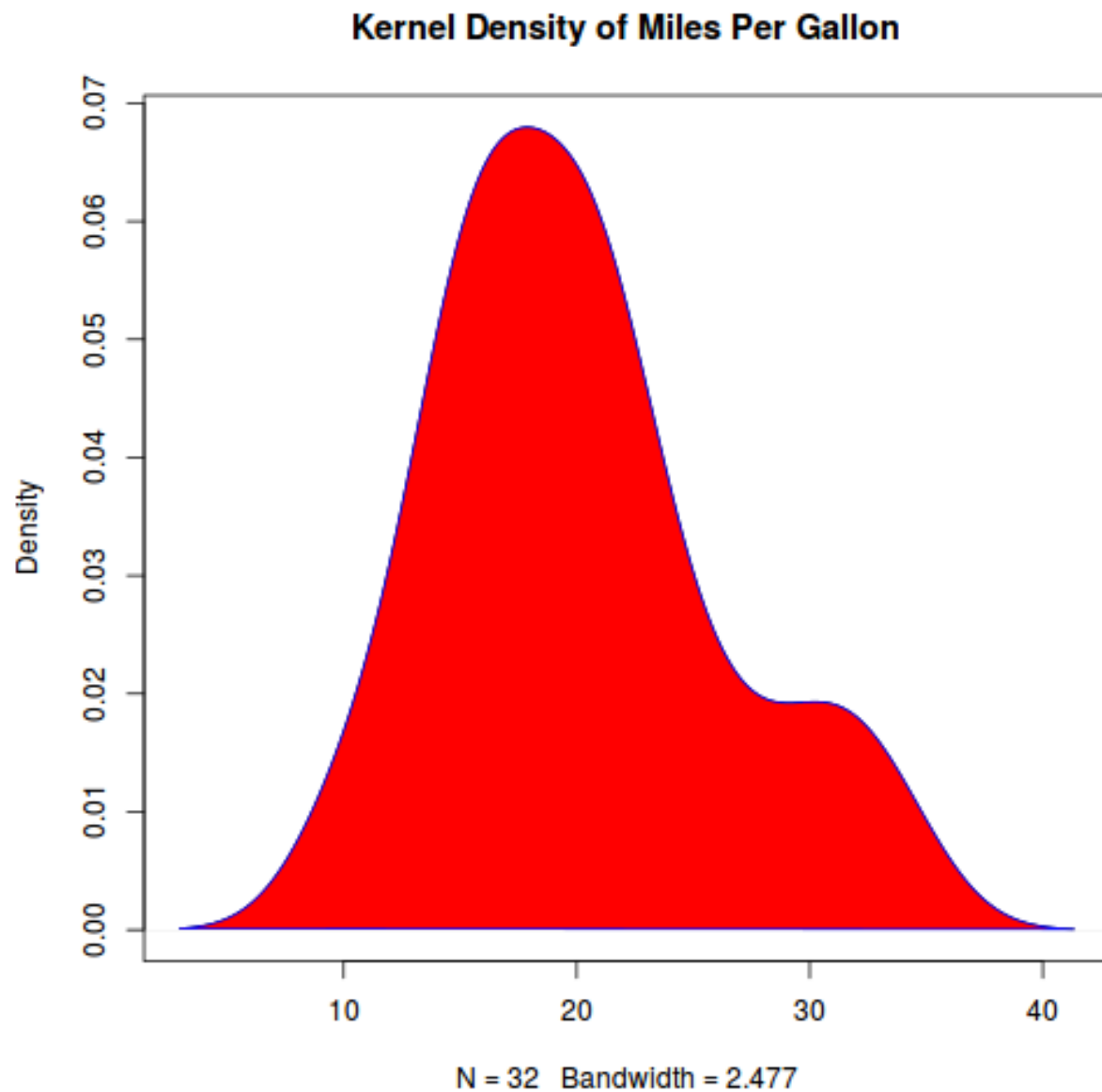
From the [Quick-R](#) website.

[Density Plots](#)

Stat. 6502 book Rice Ch. 10 Section 3

```
=====

attach(mtcars)
# Filled Density Plot
d <- density(mtcars$mpg)
plot(d, main="Kernel Density of Miles Per Gallon")
polygon(d, col="red", border="blue")
```



## Slide With R Code

From the [Quick-R](#) website.

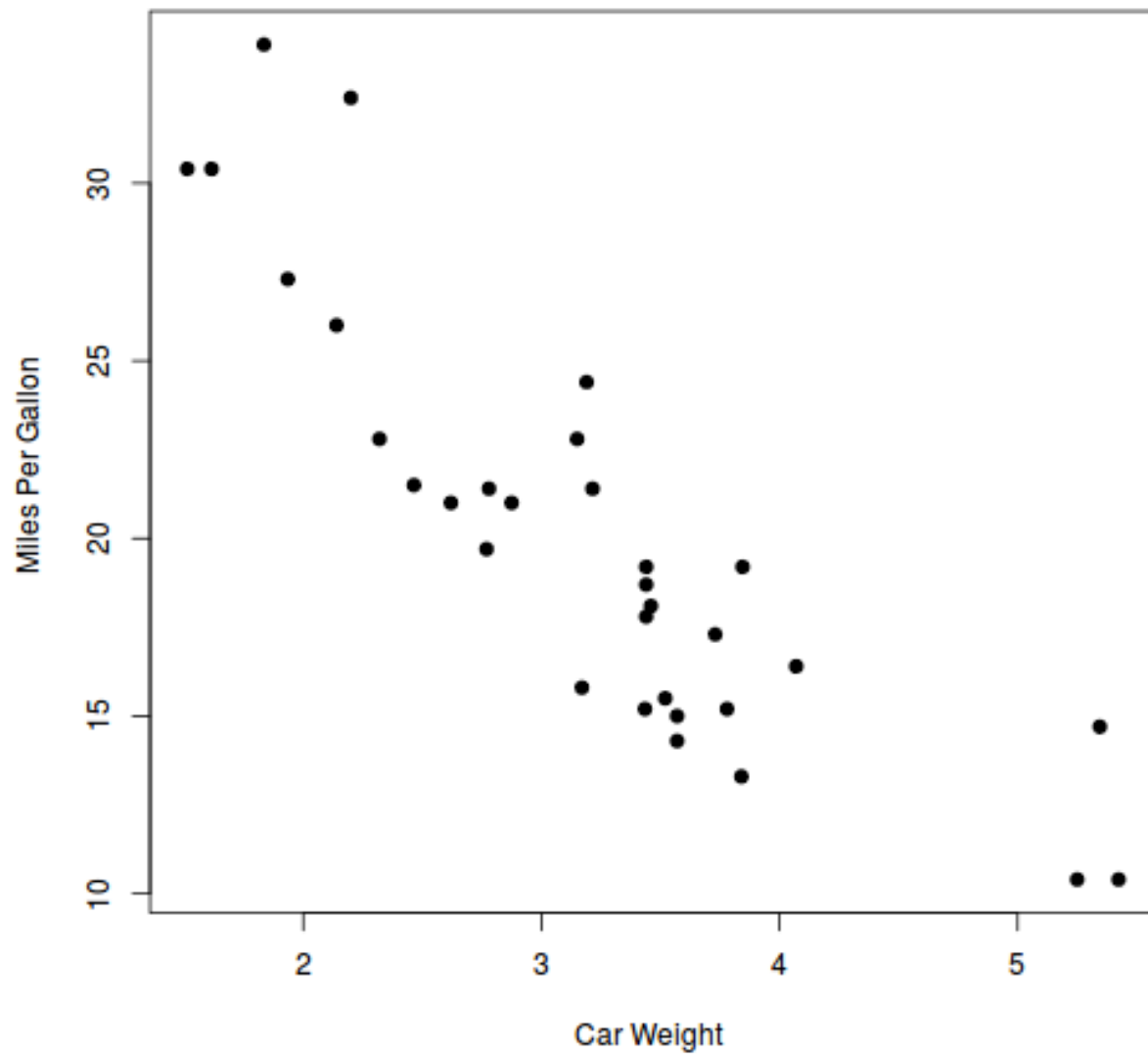
[Scatter Plots](#)

Stat. 6502 book Rice Ch. 10 Section 7

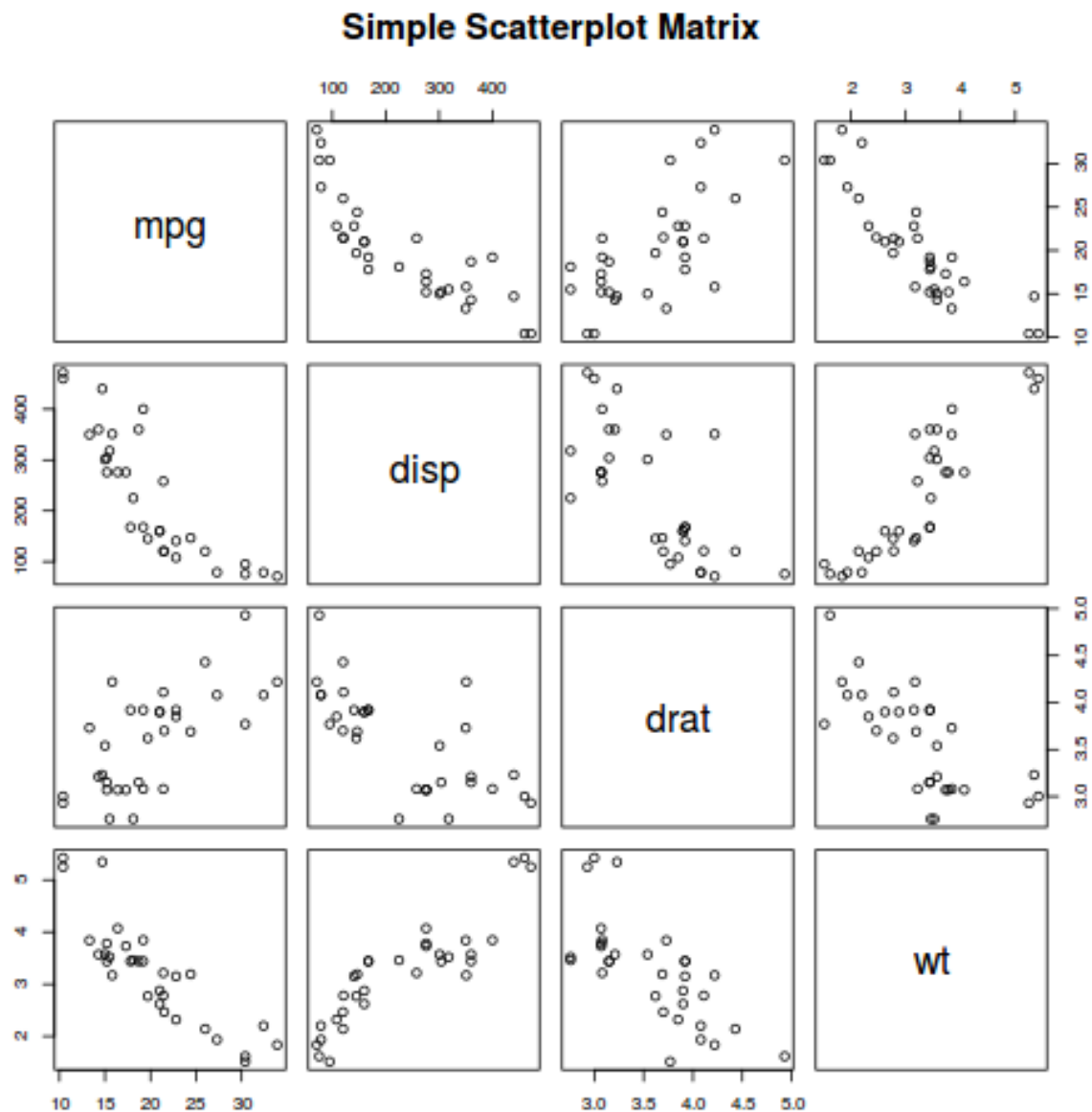
=====

```
# Simple Scatterplot
plot(wt, mpg, main="Scatterplot Example",
     xlab="Car Weight ", ylab="Miles Per Gallon ", pch=19)
```

## Scatterplot Example



```
=====
# Basic Scatterplot Matrix
pairs(~mpg+disp+drat+wt,data=mtcars,
      main="Simple Scatterplot Matrix")
```



## Slide With R Code

From the [Quick-R](#) website.

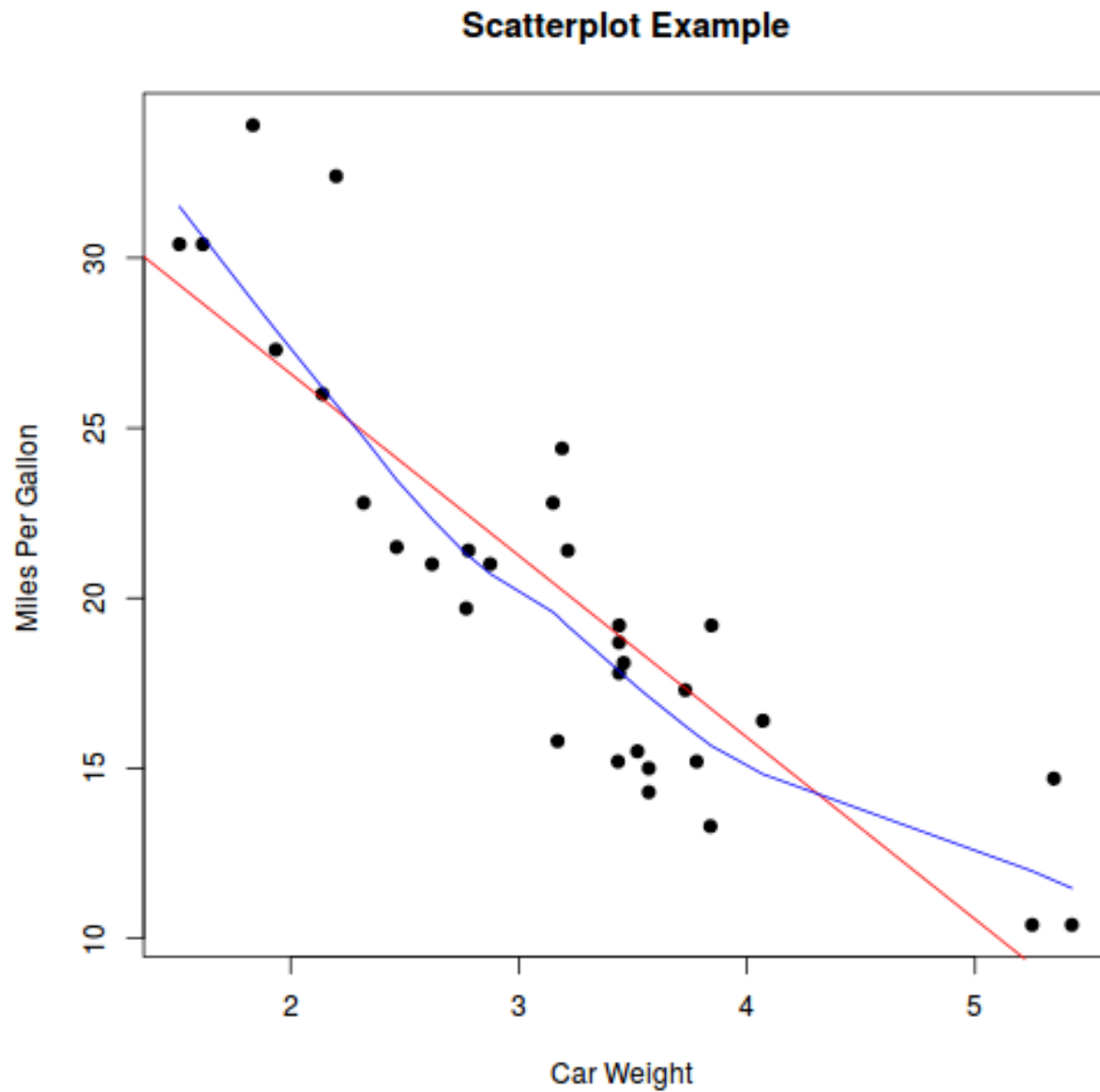
Regression line and the loess line.

Stat. 6502 book Rice Ch. 14 Section 7

=====

```
plot(wt, mpg, main="Scatterplot Example",
     xlab="Car Weight ", ylab="Miles Per Gallon ", pch=19)
# Add fit lines
```

```
abline(lm(mpg~wt), col="red") # regression line (y~x)
lines(lowess(wt,mpg), col="blue") # lowess line (x,y)
```



## Slide With R Code

From the [Quick-R](#) website.

3D Scatterplots.

=====



```
# 3D Scatterplot  
library(scatterplot3d)  
attach(mtcars)  
scatterplot3d(wt,disp,mpg, main="3D Scatterplot")
```

