

# **Data Management With R: Data Visualization**

---

Matthias Haber

4 September 2017

# Plan for Today

1. Making sure everyone is set up
2. Data visualization in base R
3. Data visualization with ggplot2
4. Excercises

## **Making sure everyone is set up**

---

# Packages

Load the tidyverse

```
library(tidyverse)
```

If you get the error message “there is no package called ‘tidyverse’”, you’ll need to first install it, then run `library()` again.

```
install.packages("tidyverse")
library(tidyverse)
```

# Data visualization in base R

---

# Why do we use graphs in data analysis?

- To understand data properties
- To find patterns in data
- To suggest modeling strategies
- To “debug” analyses
- To communicate results

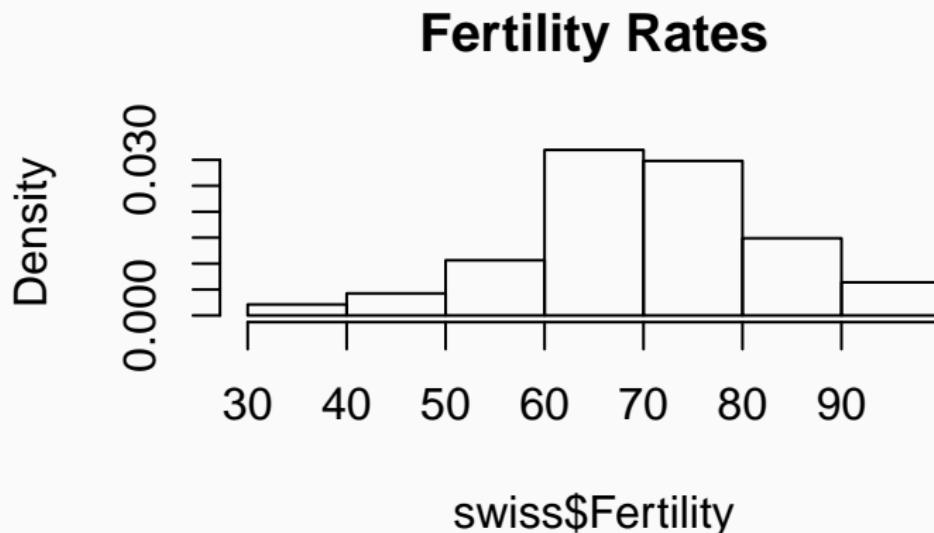
# Plotting with R

R has a powerful graphics engine to produce high quality graphs e.g.:

- `plot`: Basic plotting function (e.g. for scatter plots)
- `hist()`: Histograms
- `dotchart()`: Dot plots
- `boxplot()`: Box-and-whisker plots

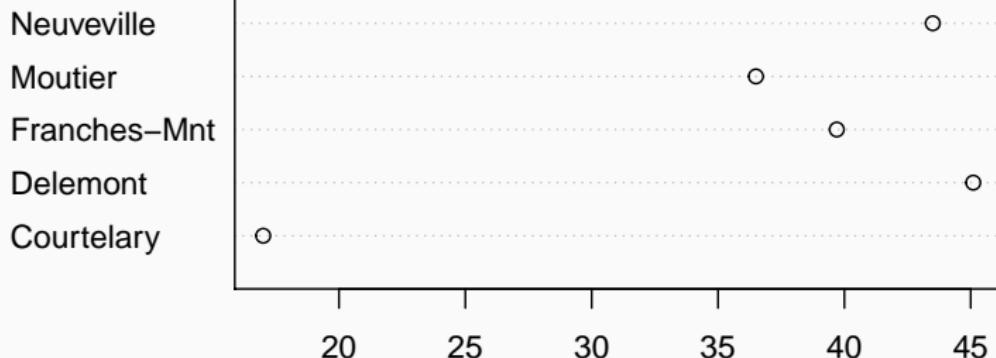
# Histograms

```
# Create a histogram  
data(swiss)  
hist(swiss$Fertility, freq=FALSE, main="Fertility Rates")
```



## Dotchart

```
# Create a dot plot  
data(swiss)  
dotchart(swiss[1:5,2], labels=rownames(swiss))
```



# Data visualization with ggplot2

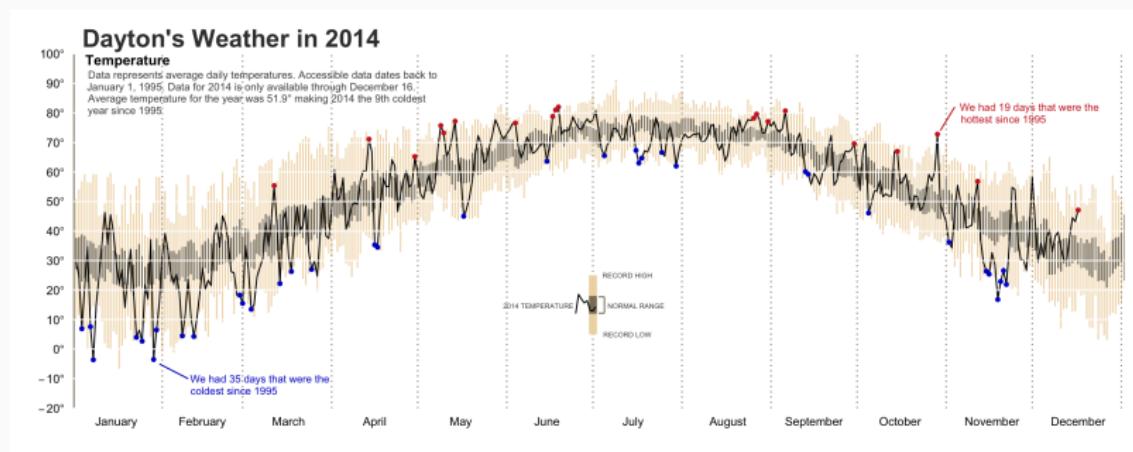
---

## Why use ggplot2

Base R functions like `plot` and `hist` are great for quickly displaying data but are not very aesthetic, and are limited in the quantity and type of information conveyed.

R has several systems for making graphs, but `ggplot2` is one of the most elegant and most versatile. `ggplot2` implements the grammar of graphics, a coherent system for describing and building graphs.

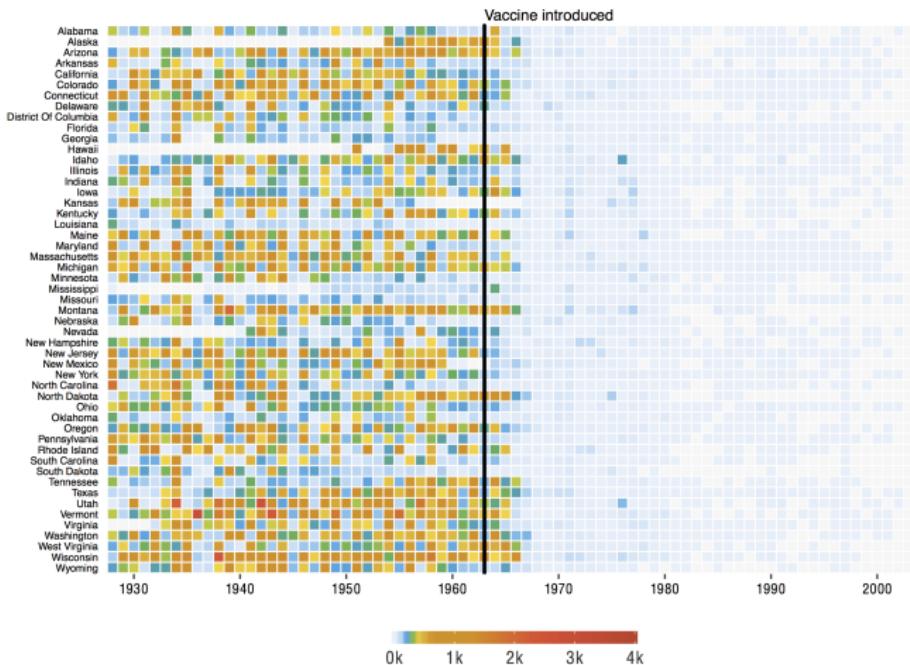
# ggplot2 examples



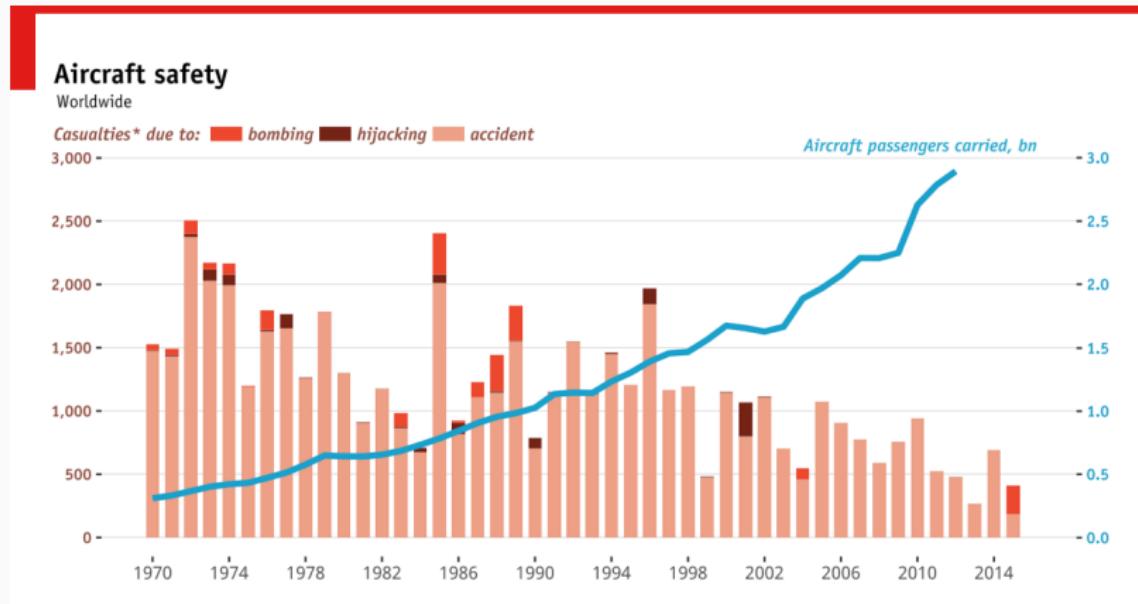
Source

# ggplot2 examples

## Measles

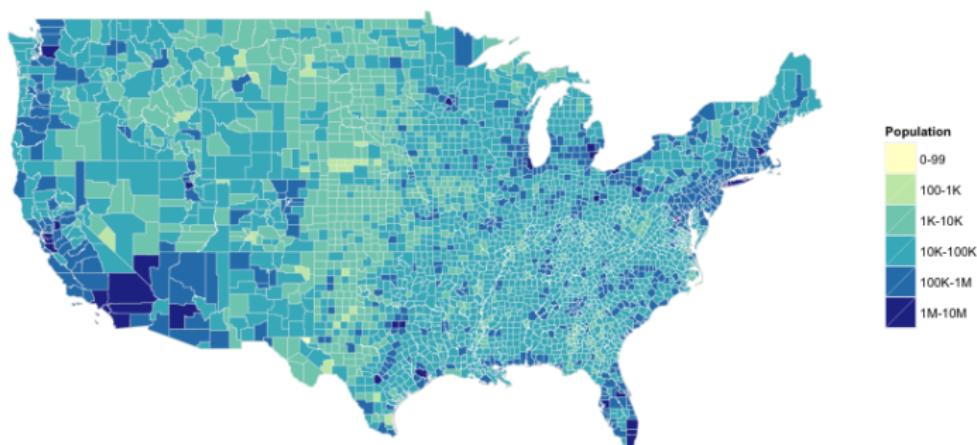


# ggplot2 examples



Source

# ggplot2 examples

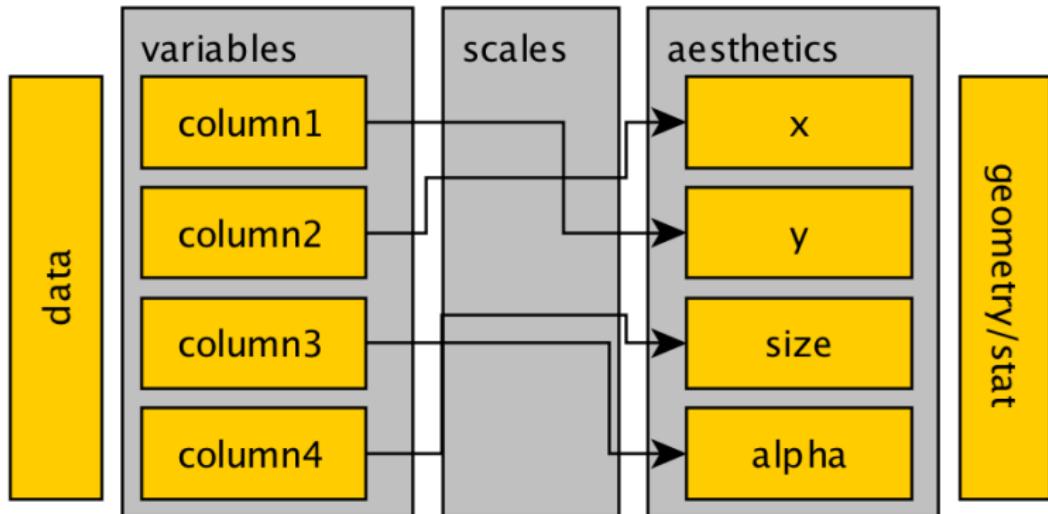


Source

# The grammar of graphics

- Each plot is made of layers. Layers include the coordinate system (x-y), points, labels, etc.
- Each layer has aesthetics (aes) including x & y, size, shape, and color.
- The main layer types are called geometrics(geom) and include lines, points, etc.

# The grammar of graphics



Source

# Components of a ggplot2 graph

- **data:** Variables mapped to aesthetic attributes
- **aesthetic:** Visual property of the plot objects
- **geom:** Geometrical object used to represent data
- **stats:** Statistical transformations of the data
- **scales:** Values mapped to aesthetic attributes
- **coord:** Coordinate system
- **facets:** Subplots that each display one subset of the data

## ggplot2 basic

A ggplot is build piece by piece

1. We need some data to be plotted

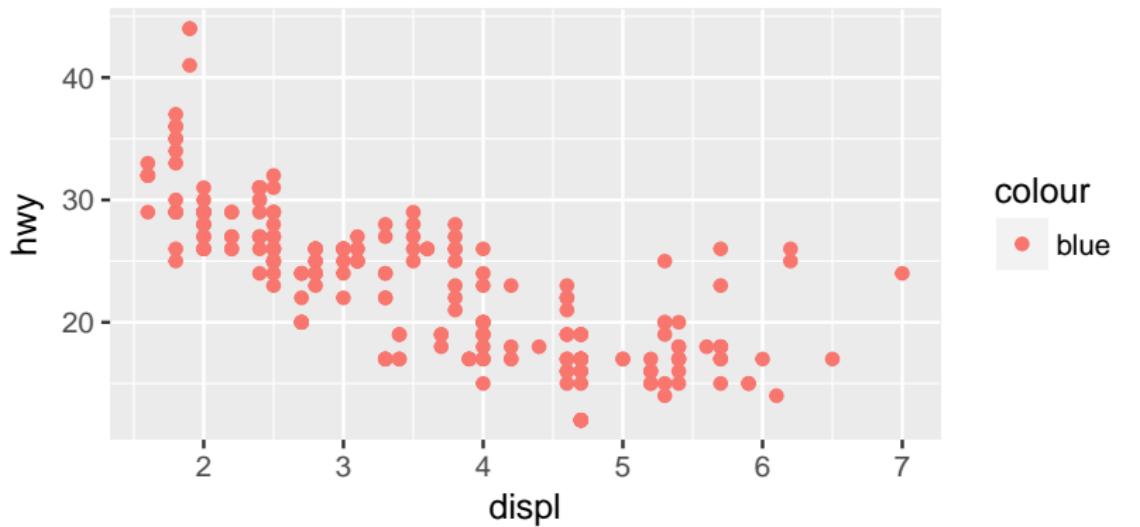
## Exercises

---

# Aesthetic mappings

1. What's gone wrong with this code?

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy,  
                           color = "blue"))
```

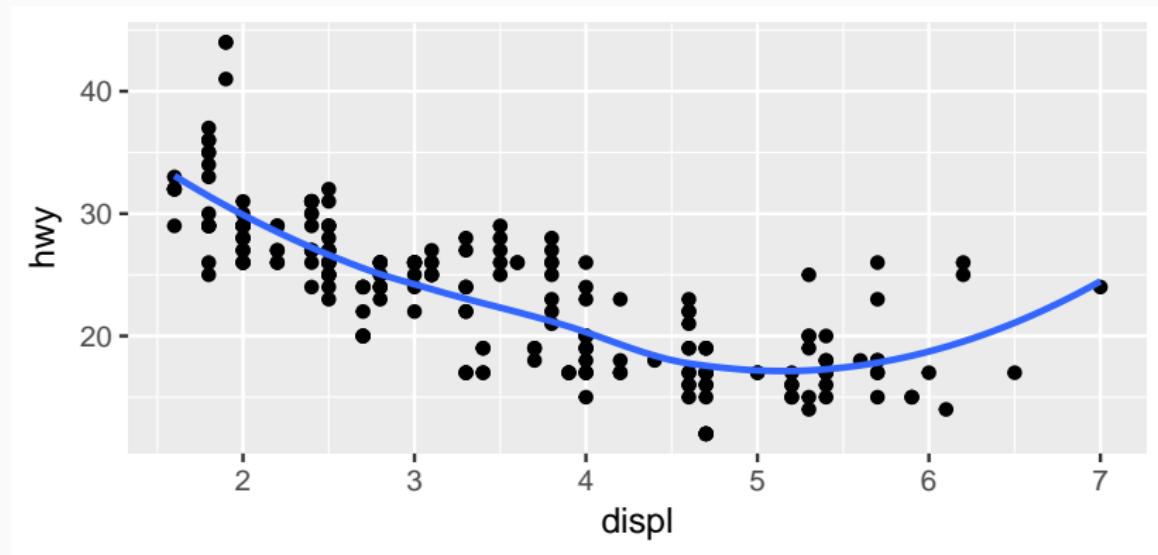


## Aesthetic mappings

2. Map a continuous variable from *mpg* to color, size, and shape.  
How do these aesthetics behave differently for categorical  
vs. continuous variables?

## Geometric objects

Recreate the R code necessary to generate the following graph.



## Group Exercises

---

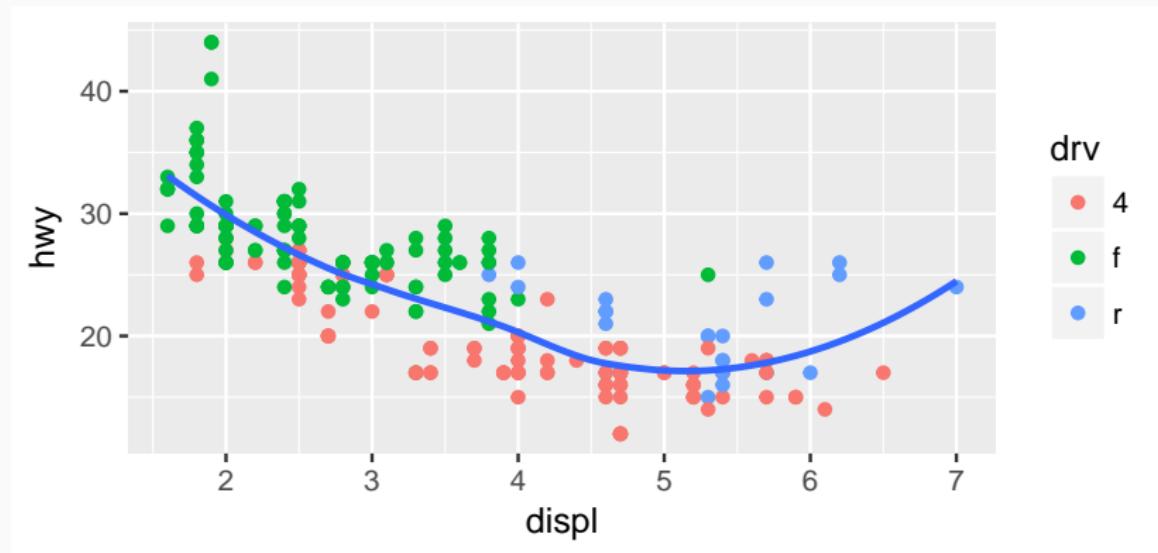
## Group Exercises

In the rest of the class, I want you to work in teams of two and try to reproduce the following plots.

If you need help, this Cheatsheet may come in handy.

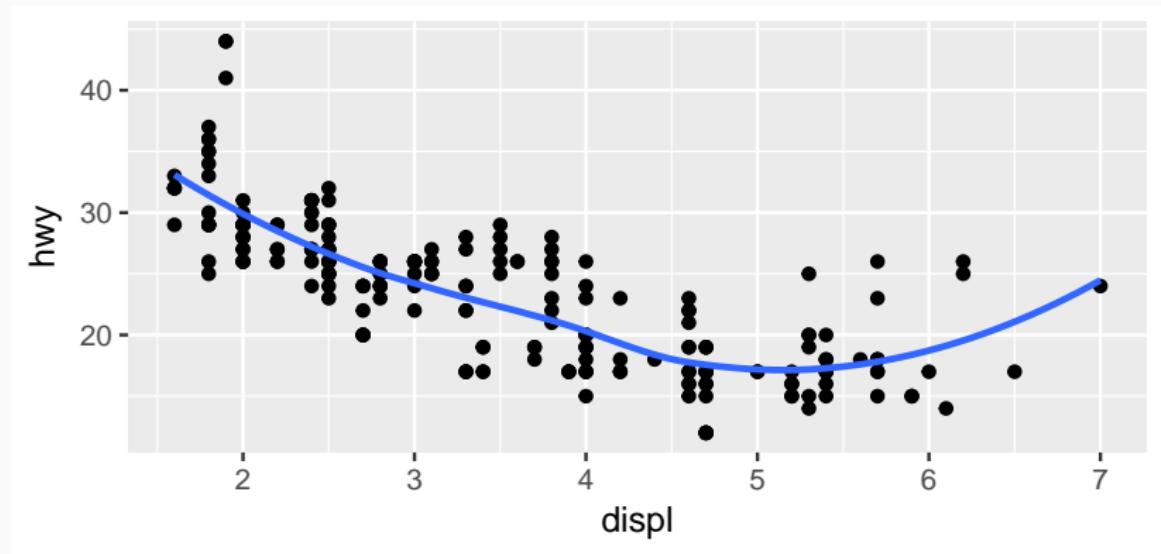
# Plot 1

Recreate the R code necessary to generate the following graph.



## Plot 2

Recreate the R code necessary to generate the following graph.



# Final thoughts



Source

**That's it for today. Questions?**