

MassMutual DSDP 2018:

VISUALIZATION TECHNIQUES

June 18, 2018

R. Jordan Crouser

Assistant Professor of Computer Science

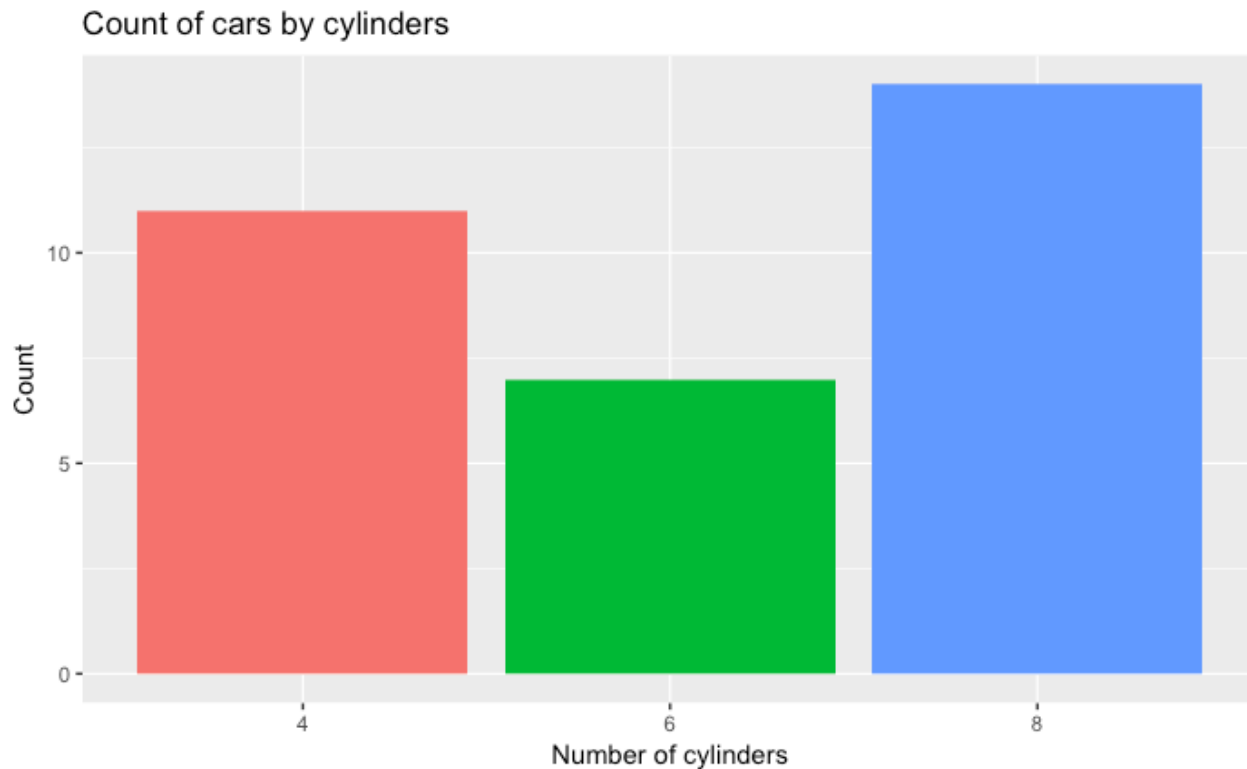
Smith College

What visualization techniques do you know?



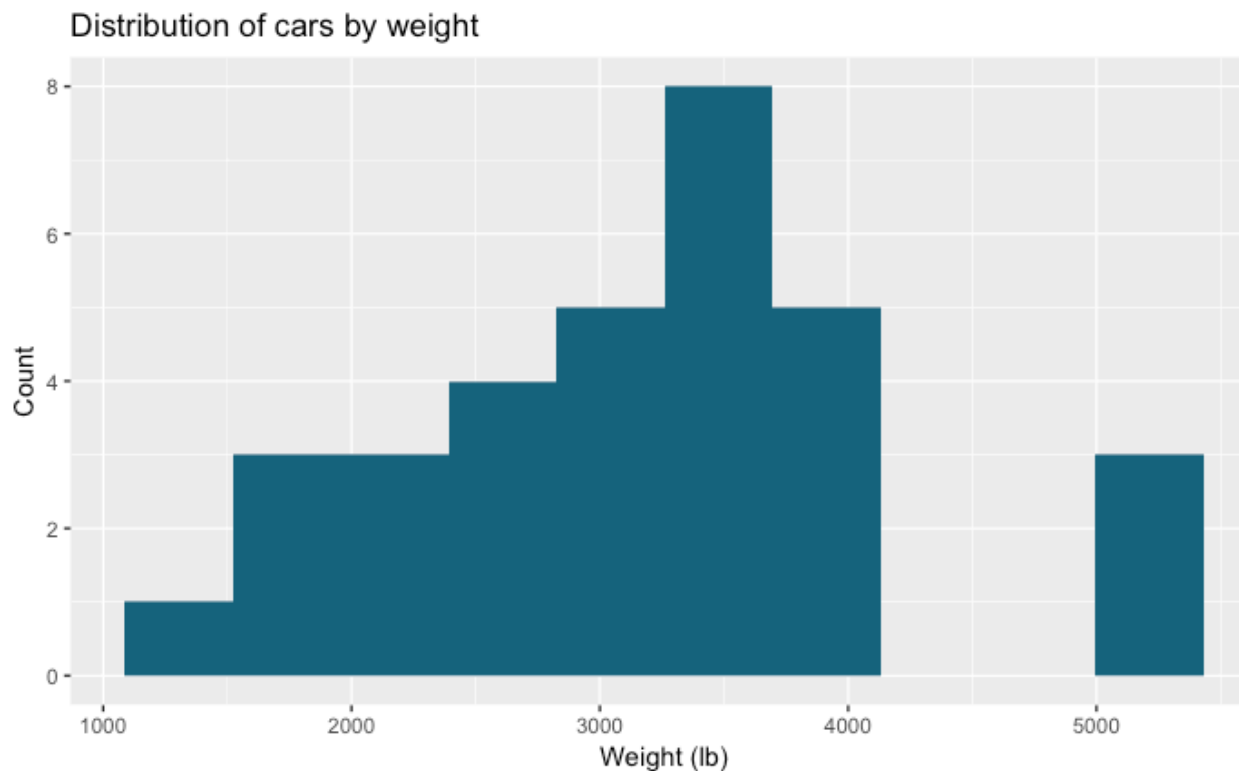
Bar chart

- Used for **comparable variables**
- Compares **quantitative** values for different categories
- Highlights **relative amounts**
- Grouped/stacked bars can break each **category** into sub-groups



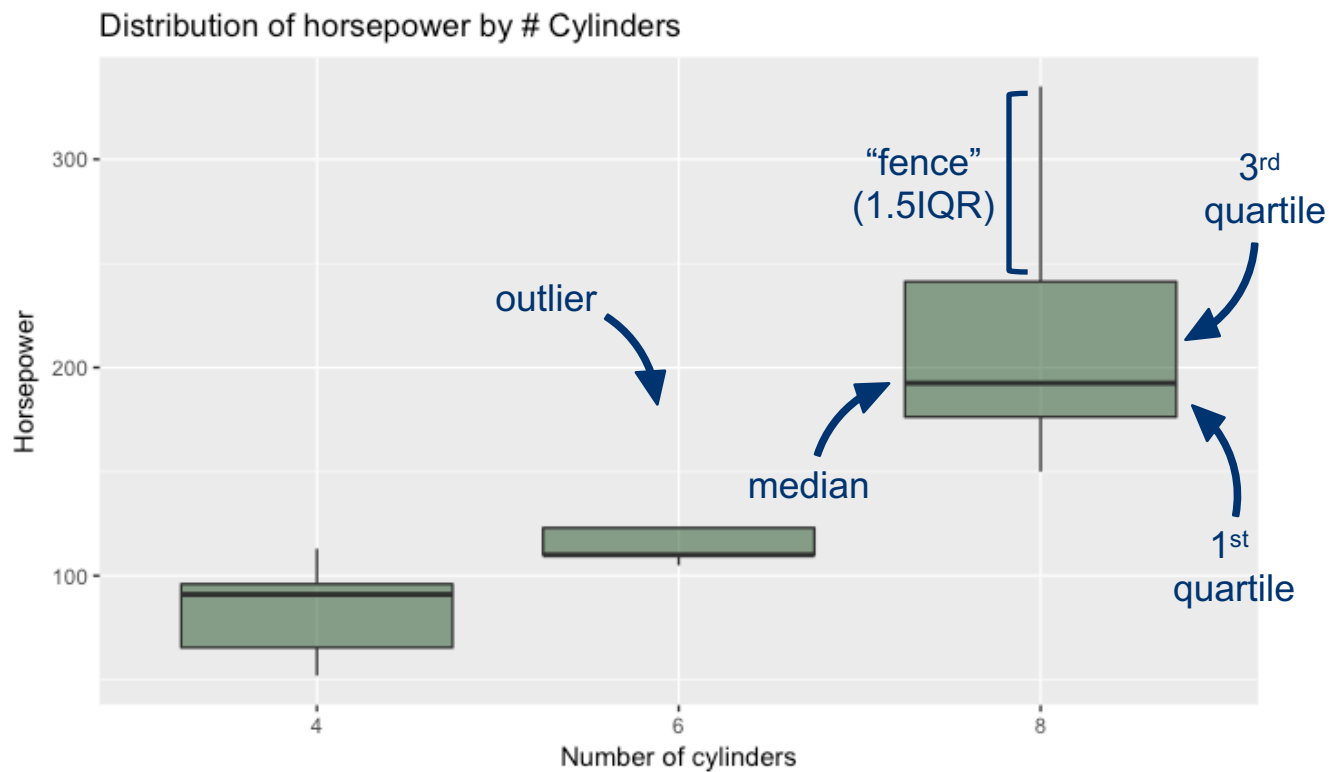
Histogram

- Looks like a bar chart... but the x-axis is **continuous**
- Y-axis shows count or relative frequency
- Highlights **distribution**
- Note: bin size makes a big difference!



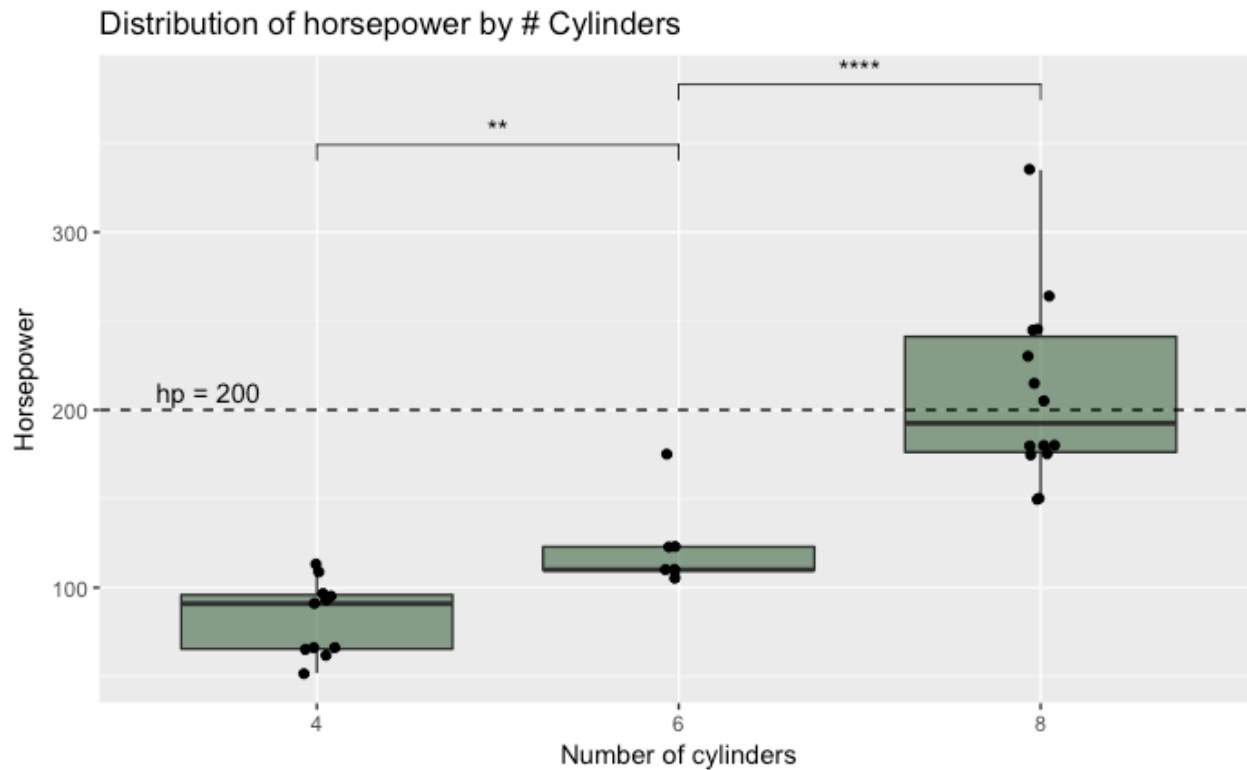
Boxplot

- Also useful for highlighting **distribution**
- Calls out key values:
 - median
 - 1st & 3rd quartiles
 - “fences”
 - outliers



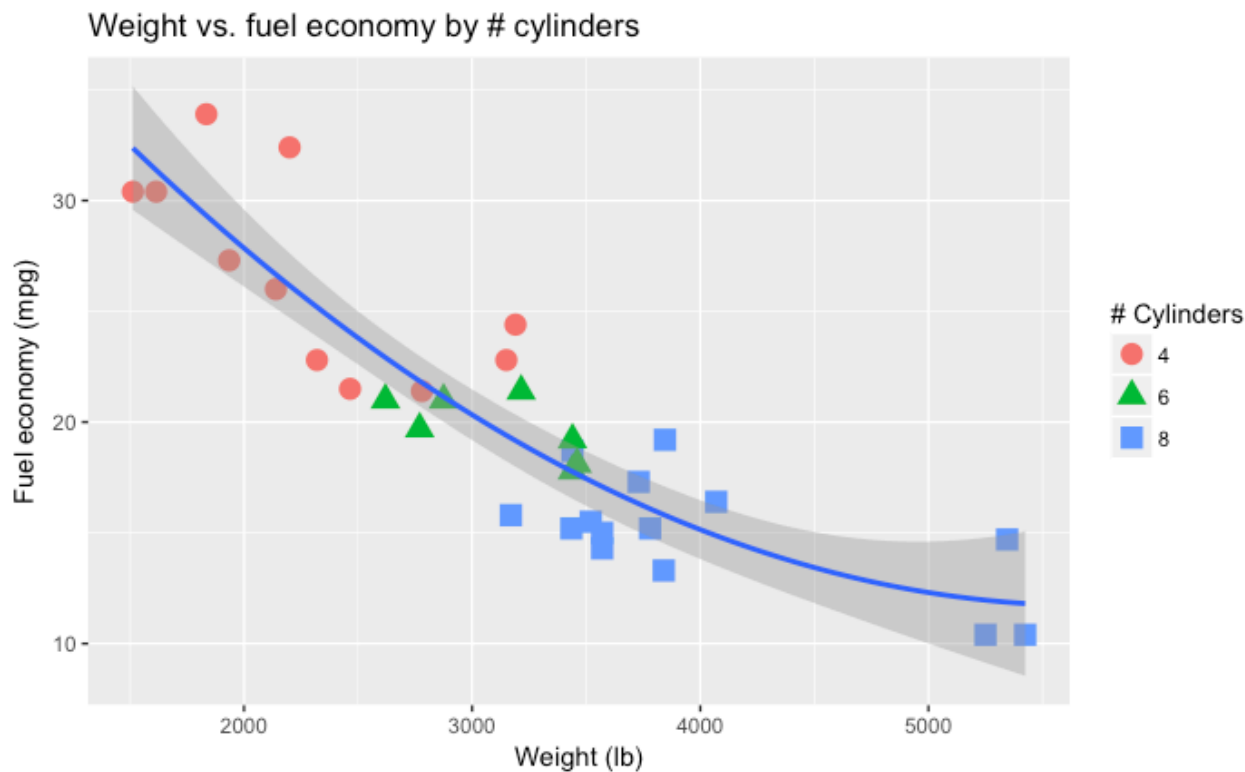
Boxplot add-ons

- Use “jitter” to show actual values
- Reference lines can help provide context
- Can use annotations to show statistical significance



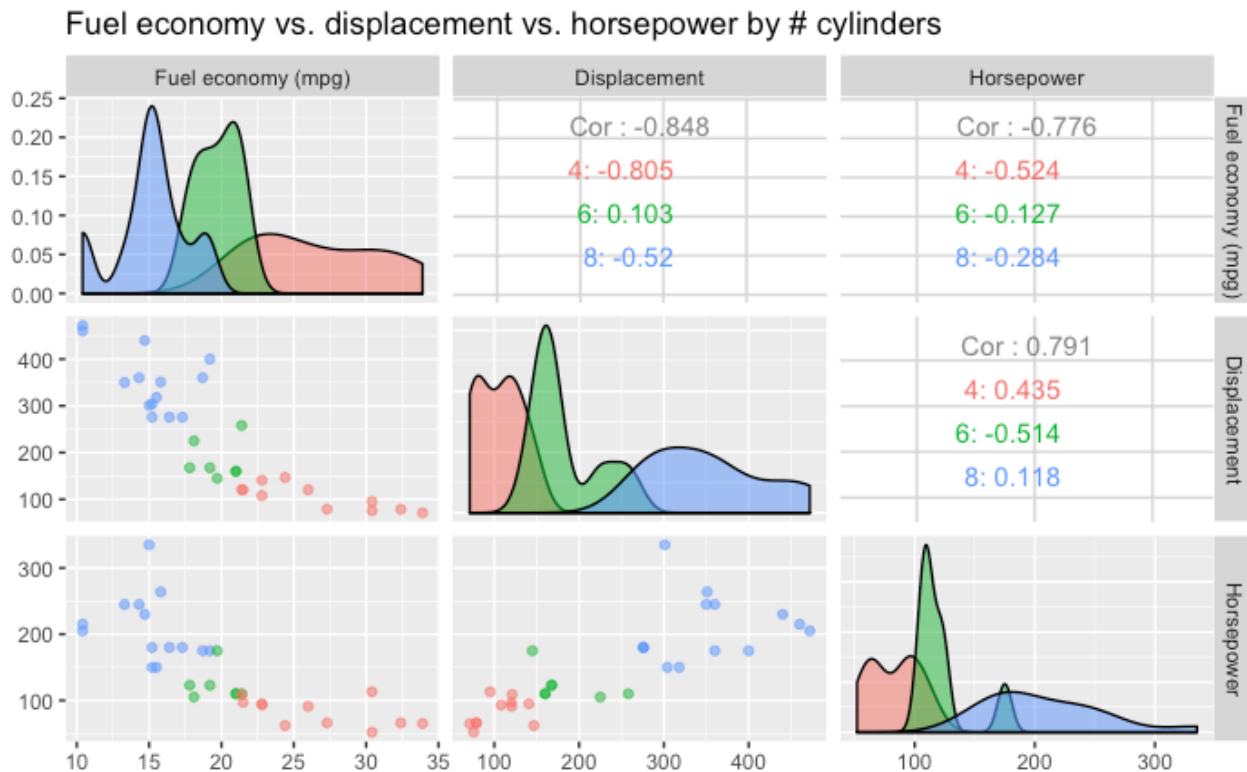
Scatterplot

- Shows the relationship between two **continuous variables**
- Each point in the plot represents an observation
- You can change color or symbol to **highlight groups**
- Sometimes useful to show a trend line (regression)



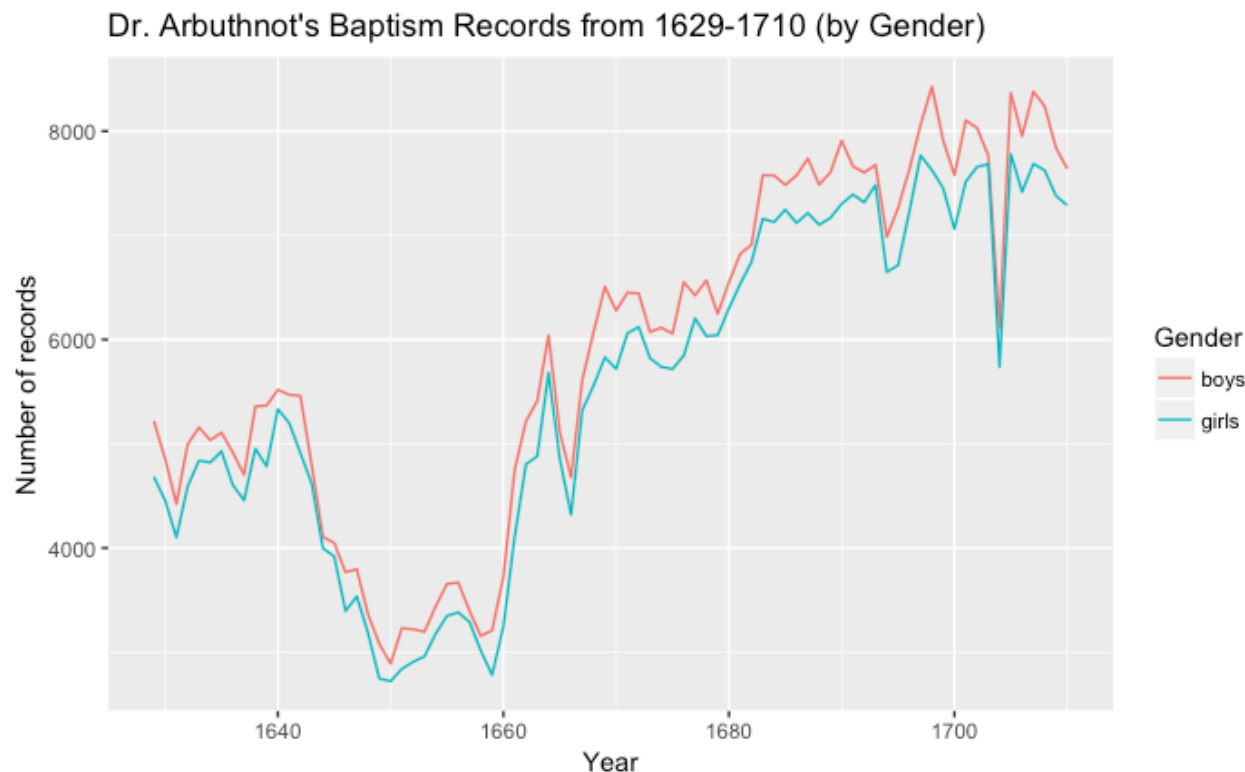
Scatterplot matrix (SPLOM)

- Scatterplots show the relationship between just two **continuous variables at a time**
- We can combine multiple scatterplots into a matrix to show additional relationships



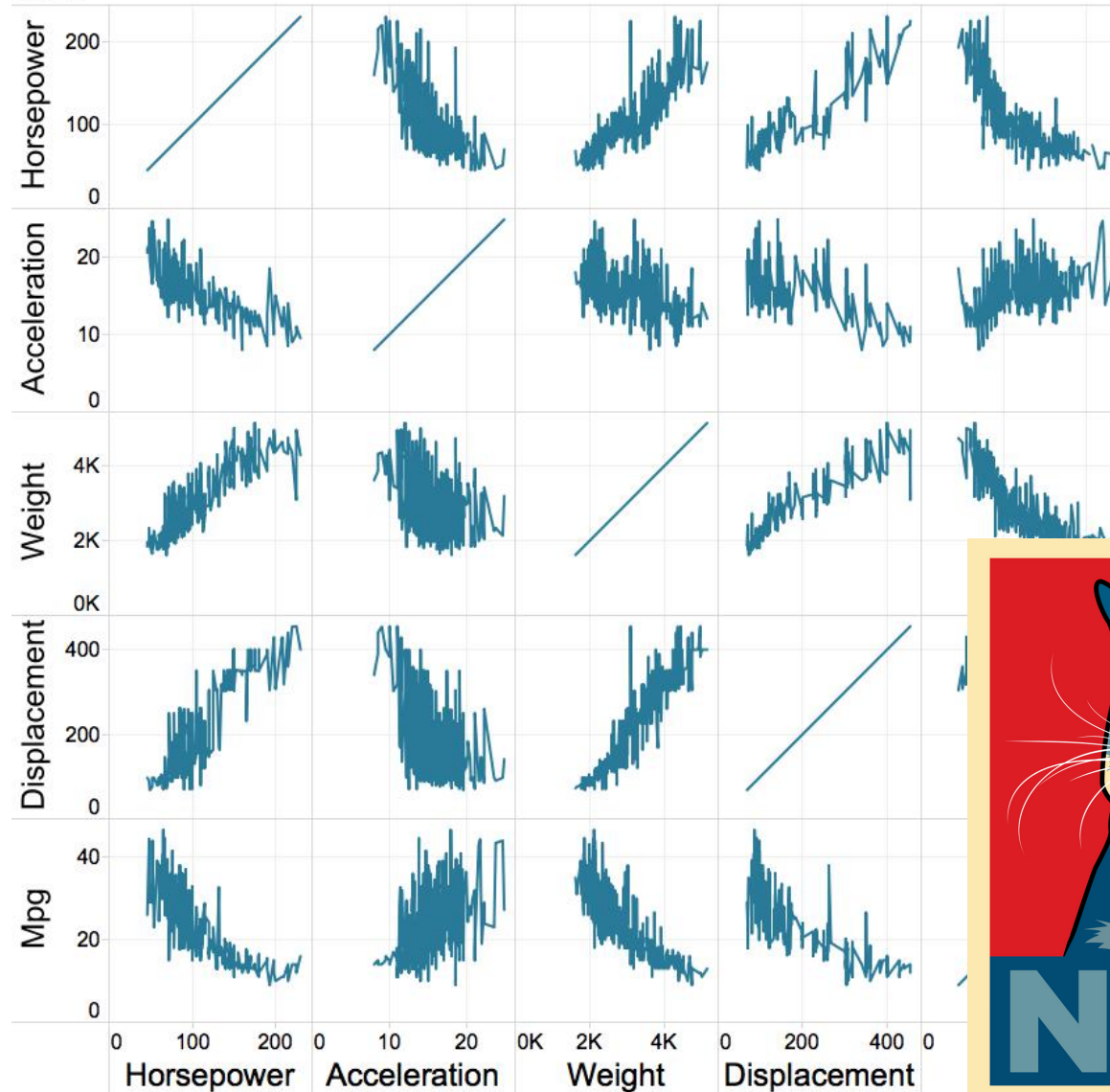
Line chart

- Shows the trend in one variable, often **over time**
- Multiple lines can show multiple variables, or the same variable for multiple observations (must have the **same scale!**)
- Highlights “position switches”

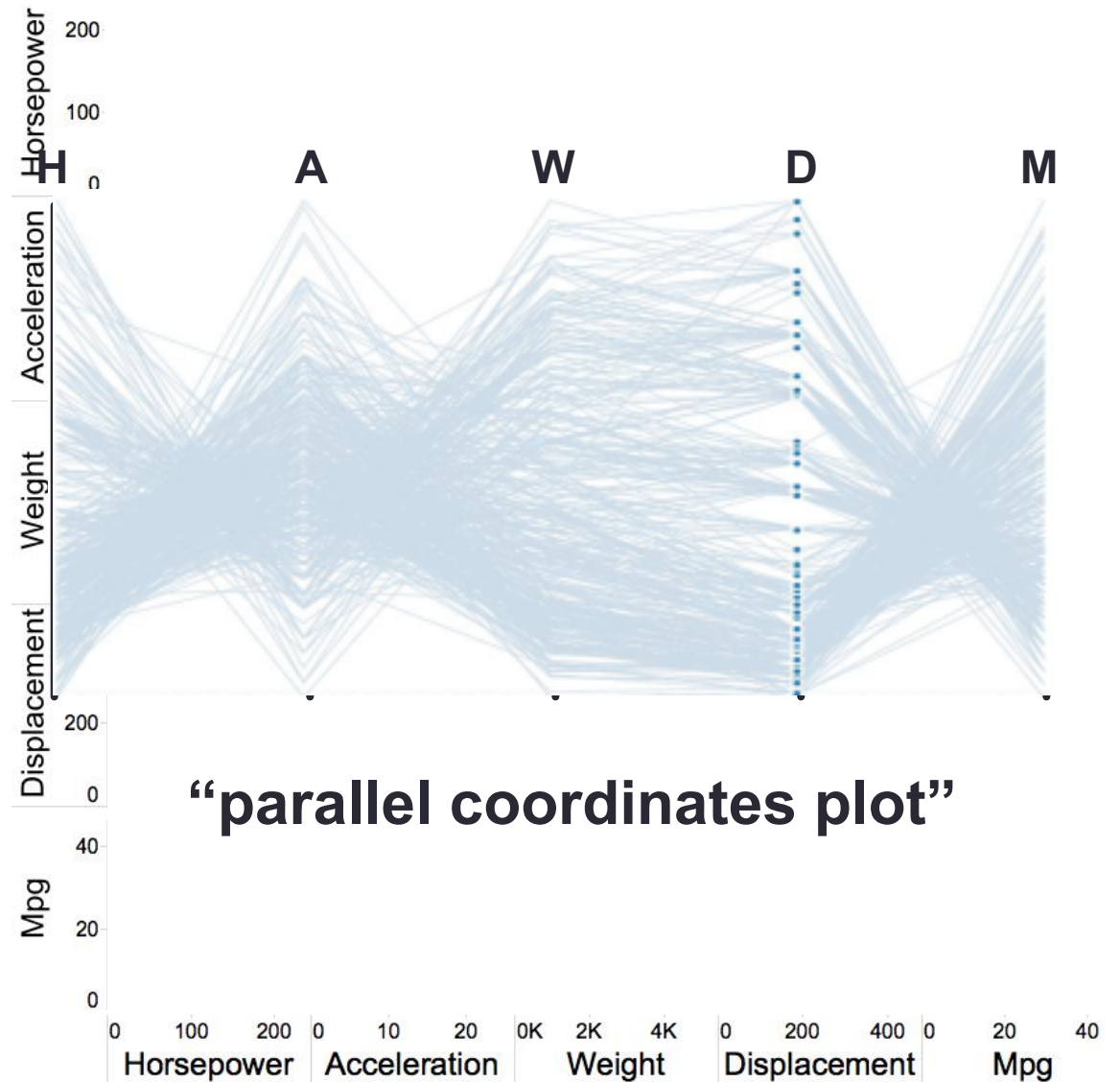


Multiple variables: line chart matrix?

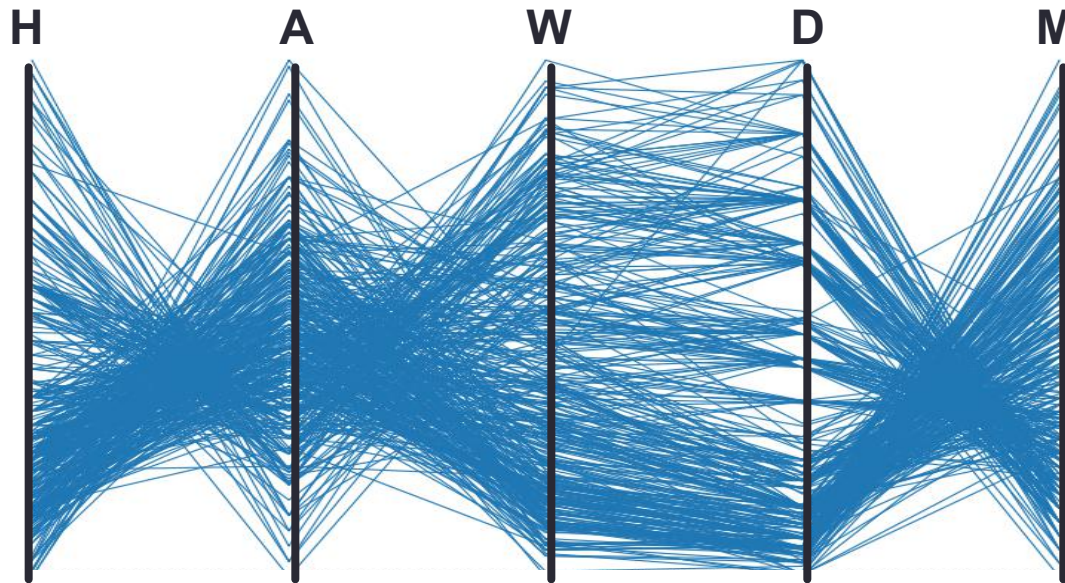
Sheet 3



Weirder idea



Morning challenge



“parallel coordinates plot”

How would you build
this using `ggplot2`?

