

# Lecture 3: Continuous variables

Michael Lopez, Skidmore College

# Goals: charts for continuous data

1. Univariate: histogram, boxplot, violin, density curve
2. Bivariate: scatter, line
3. Tricks: jittering, grouping, coloring

# Data set

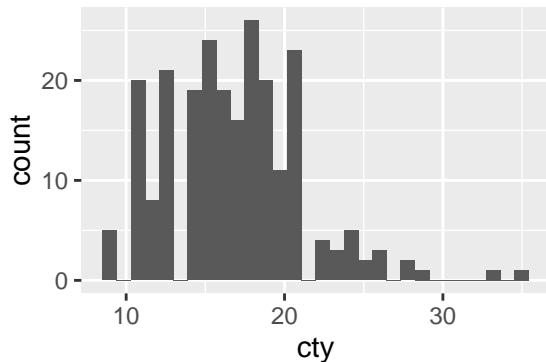
```
library(ggplot2); library(dplyr)
mpg %>% head(4)
```

```
## # A tibble: 4 × 11
##   manufacturer model displ  year  cyl    trans  drv   cty   hwy fl
##       <chr>   <chr> <dbl> <int> <int>    <chr> <chr> <int> <int> <chr>
## 1      audi    a4    1.8  1999     4 auto(l5)   f    18    29   p
## 2      audi    a4    1.8  1999     4 manual(m5) f    21    29   p
## 3      audi    a4    2.0  2008     4 manual(m6) f    20    31   p
## 4      audi    a4    2.0  2008     4 auto(av)   f    21    30   p
## # ... with 1 more variables: class <chr>
```

Variable types?

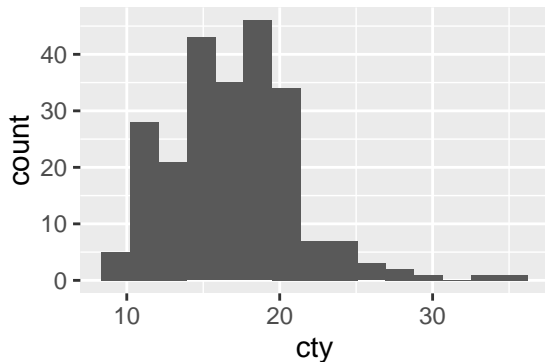
# Histogram

```
ggplot(data = mpg, aes(x = cty)) +  
  geom_histogram()
```



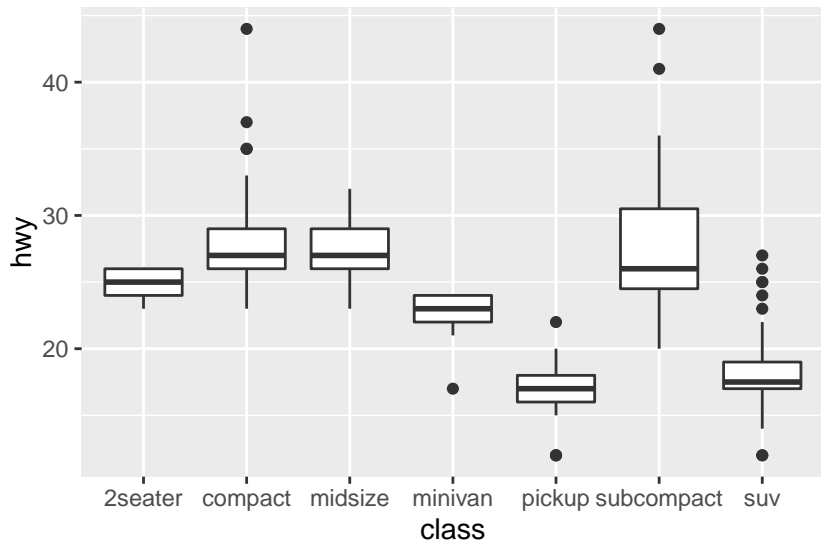
# Histogram

```
ggplot(data = mpg, aes(x = cty)) +  
  geom_histogram(bins = 15)
```



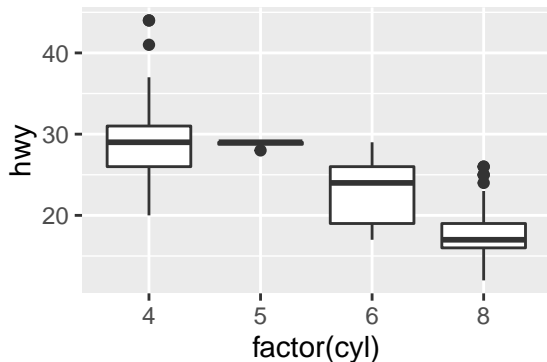
# Boxplot

```
ggplot(mpg, aes(x = class, y = hwy)) +  
  geom_boxplot()
```



## Boxplot with factor()

```
ggplot(mpg, aes(x = factor(cyl), y = hwy)) +  
  geom_boxplot()
```

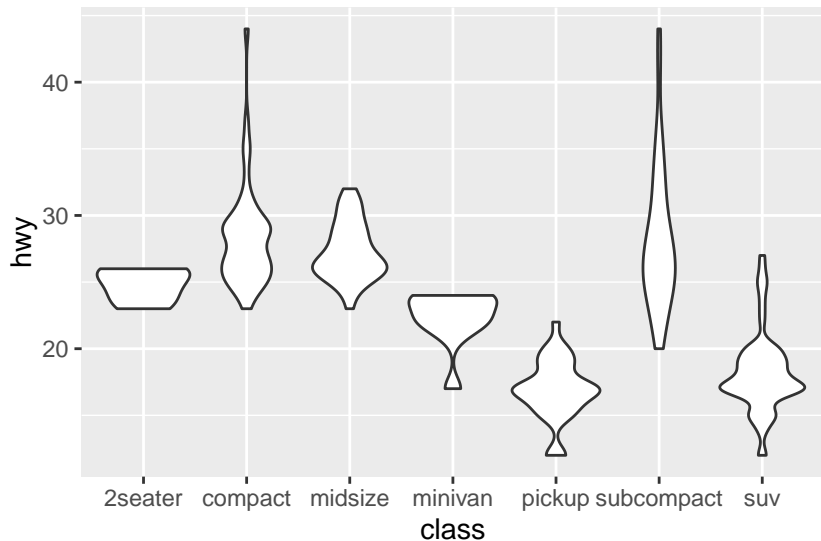


# Characteristics of a boxplot



## Violin plot

```
ggplot(mpg, aes(x = class, y = hwy)) +  
  geom_violin()
```



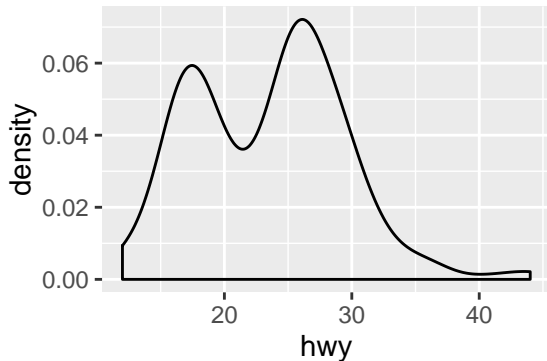
# Violin plot

```
ggplot(mpg, aes(x = class, y = hwy)) +  
  geom_violin() +  
  coord_flip()
```



# Density curves

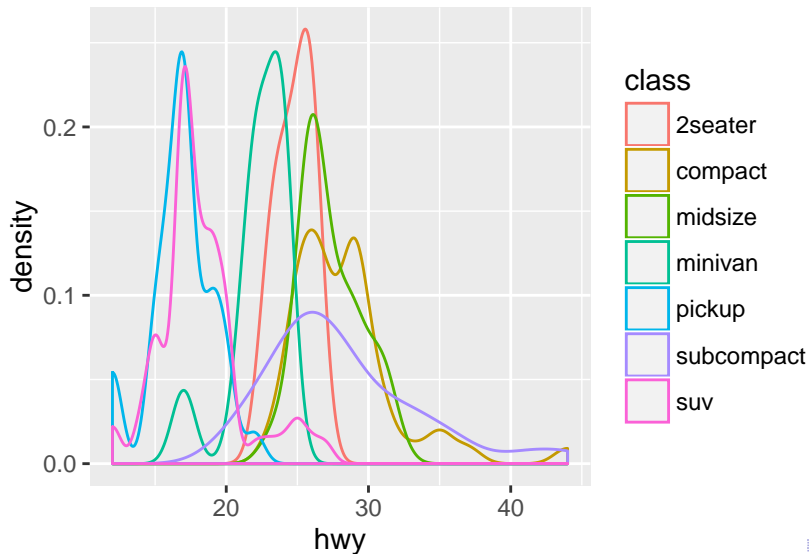
```
ggplot(mpg, aes(x = hwy)) +  
  geom_density()
```



# Anatomy of a density curve

# Density curves within groups

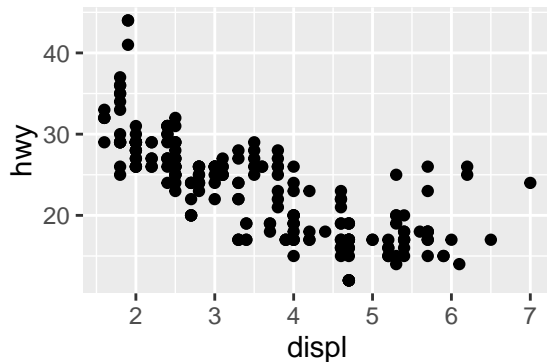
```
ggplot(mpg, aes(x = hwy, colour = class)) +  
  geom_density()
```



# Summary, univariate approaches

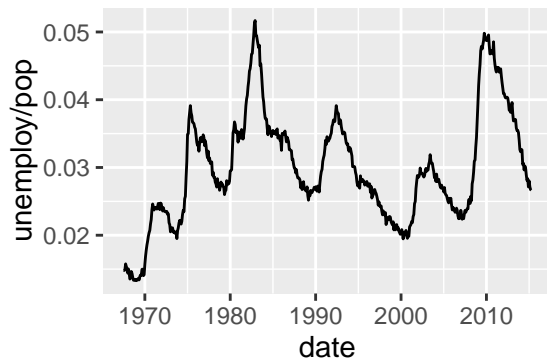
# Bivariate approaches

```
ggplot(mpg, aes(x = displ, y = hwy)) +  
  geom_point()
```



# Connections over time

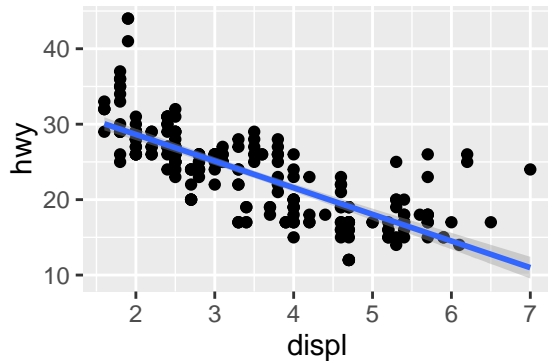
```
ggplot(economics, aes(date, unemploy / pop)) +  
  geom_line()
```





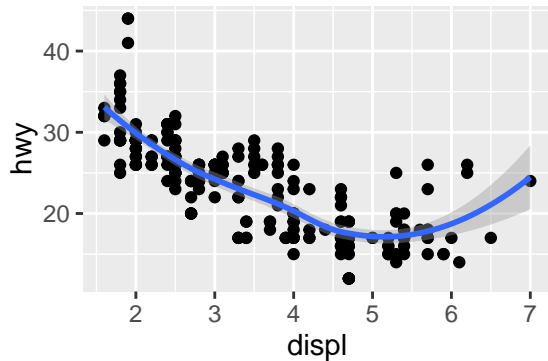
## Smoothed lines of best fit

```
ggplot(mpg, aes(x = displ, y = hwy)) +  
  geom_point() +  
  geom_smooth(method = "lm")
```



## Smoothed curves of best fit

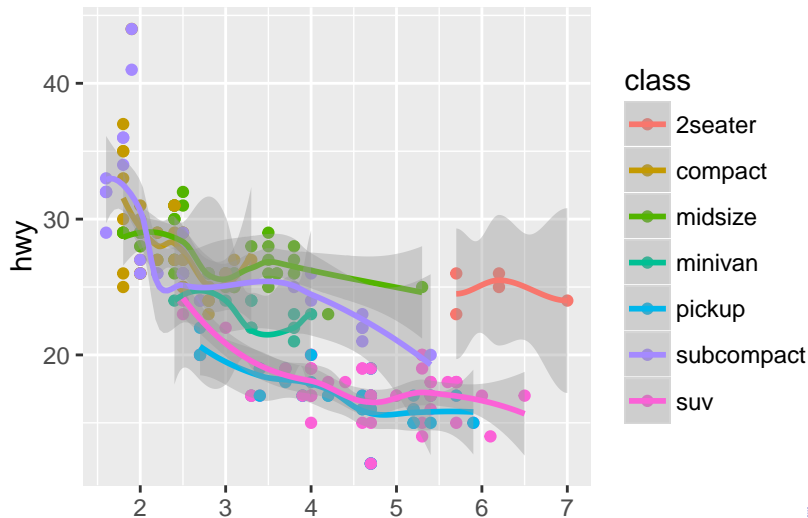
```
ggplot(mpg, aes(x = displ, y = hwy)) +  
  geom_point() +  
  geom_smooth(method = "loess")
```



# Curves versus lines of best fit

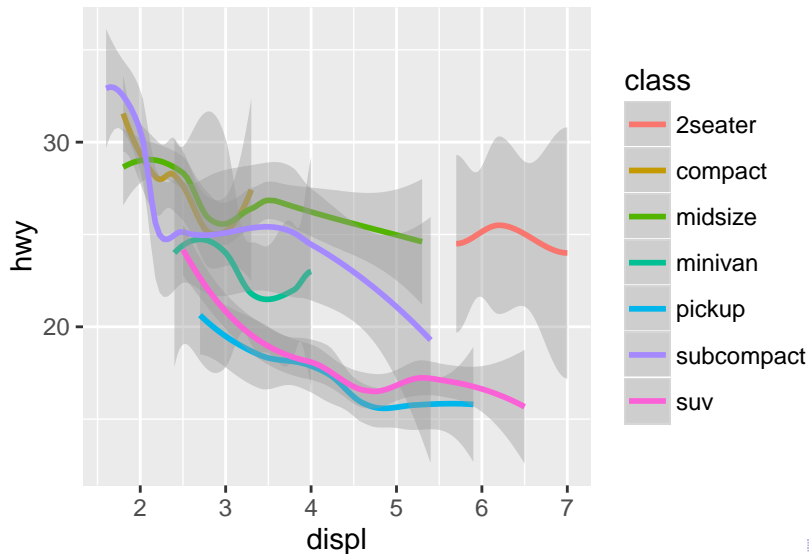
## Grouped curves of best fit

```
ggplot(mpg, aes(x = displ, y = hwy, colour = class)) +  
  geom_point() +  
  geom_smooth(method = "loess")
```



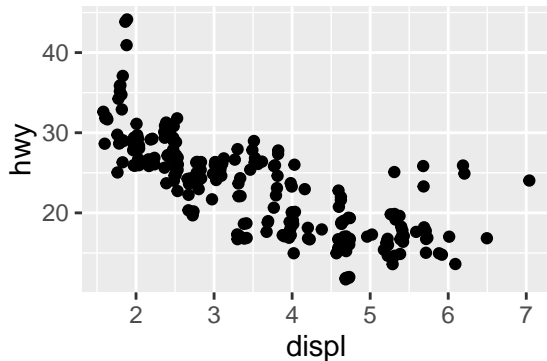
## Grouped curves of best fit

```
ggplot(mpg, aes(x = displ, y = hwy, colour = class)) +  
  geom_smooth(method = "loess")
```

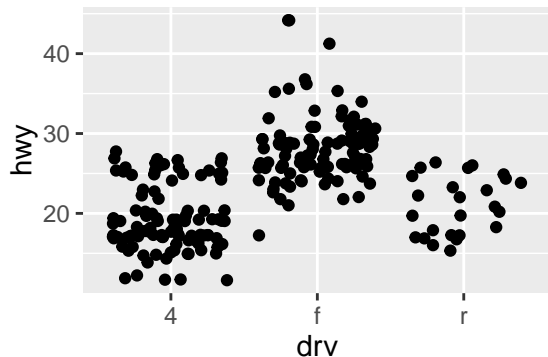


## Plotting tricks: the jitter

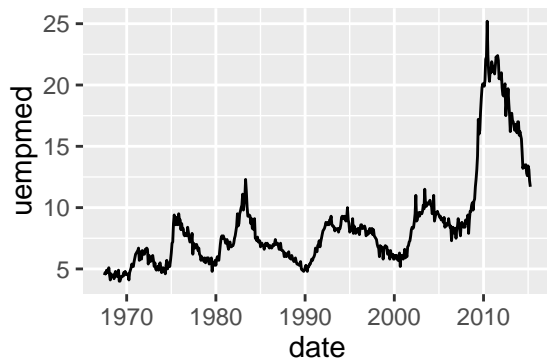
```
ggplot(mpg, aes(x = displ, y = hwy)) +  
  geom_jitter()
```



Code this plot & identify weaknesses



Code this plot & identify weaknesses

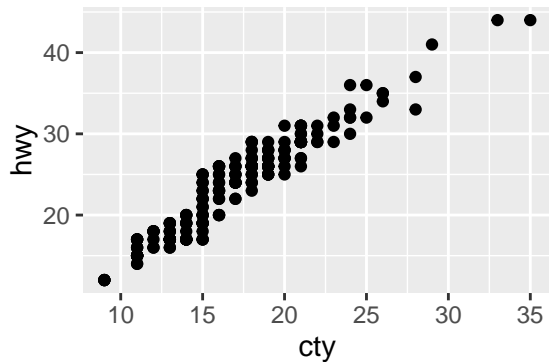




Code this plot & identify weaknesses



Code this plot & identify weaknesses



# Examples in the news

```
http://fivethirtyeight.com/features/the-most-conservative-and-most-liberal-elite-law-schools/  
http://fivethirtyeight.com/features/the-10-types-of-nfl-quarterbacks/  
http://fivethirtyeight.com/features/what-our-nba-projections-got-right-and-wrong-last-season/  
http://minimaxir.com/2015/02/ggplot-tutorial/  
http://www.nytimes.com/2016/07/20/upshot/hillary-clinton-has-a-76-percent-chance-to-win-the-presidency.html?_r=0  
http://fivethirtyeight.com/features/our-47-weirdest-charts-from-2015/  
http://fivethirtyeight.com/features/the-seven-ways-to-become-an-animation-powerhouse/  
http://fivethirtyeight.com/features/tornado-season-is-off-to-a-slow-start-but-theres-no-predicting-whats-next/  
http://fivethirtyeight.com/features/lionel-messi-is-impossible/
```