PSY 3307

Central Tendency

Jonathan A. Pedroza, MS, MA

Cal Poly Pomona

2021-08-26

Agenda

- Greek Terms
- Central Tendency
- Measures of Central Tendency
- The Mode
- The Median
- The Mean
- When to Use Each Measure
- Using The Mean in Research
  - Deviation
- Describing Population Mean

Greek Terms

$$\Sigma = Sum of scores$$

```
21+ 21+ 25+ 23+ 24+ 23+ 20+ 19+ 24+ 25+ 21+ 21+ 26+ 22+ 22+ 24+ 20+ 19+
```

```
## [1] 696
```

$$\Sigma\ X = Sum\ Of\ X$$

```
x <- c(21, 21, 25, 23, 24, 23, 20, 19, 24, 25, 21, 21, 26, 22, 22, 24, 2
sum(x)
```

```
## [1] 696
```

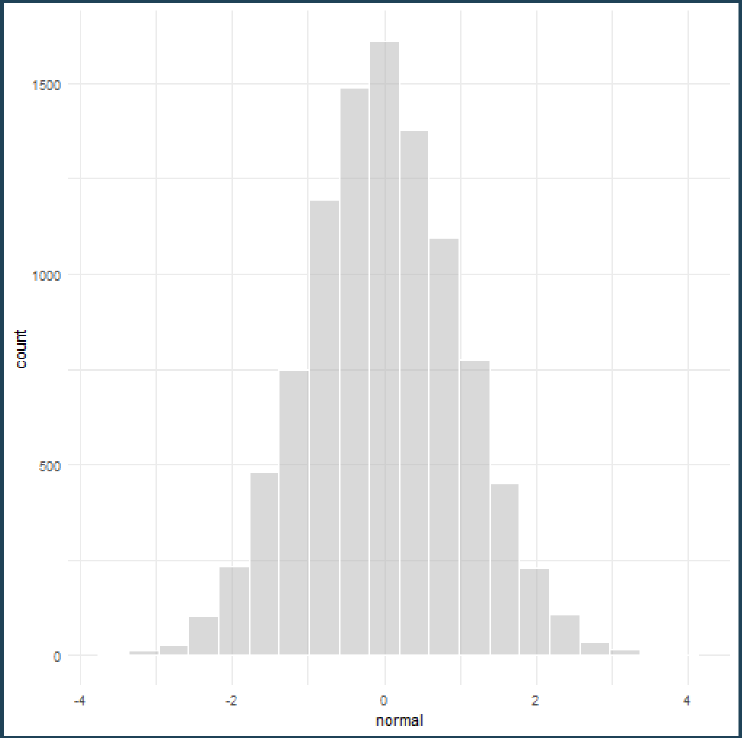
- Both of these are stating that we are adding all of the data points together.
- **Sum of X** is the sum of the scores in a sample
  - X is just another way to say all the data points for the variable.

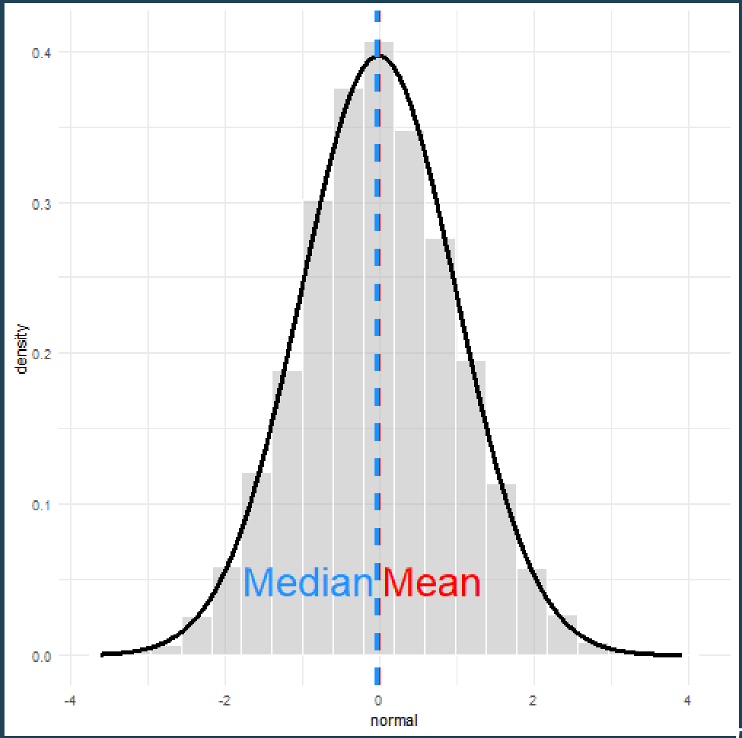
Central Tendency

- Concept that as statisticians every person is just a data point
- We are interested in the central score
  - We are interested in how much a person is away from that central score
  - But when it comes to statistics we like to group together our participants' scores/values
- **Measures of Central Tendency** are statistics that summarize the location of a distribution on a variable by indicating where the center is
- A Normal distribution will have the central point right down the center
- A skewed distribution will have the central point where the frequency of scores is the highest

Normal Distribution Example

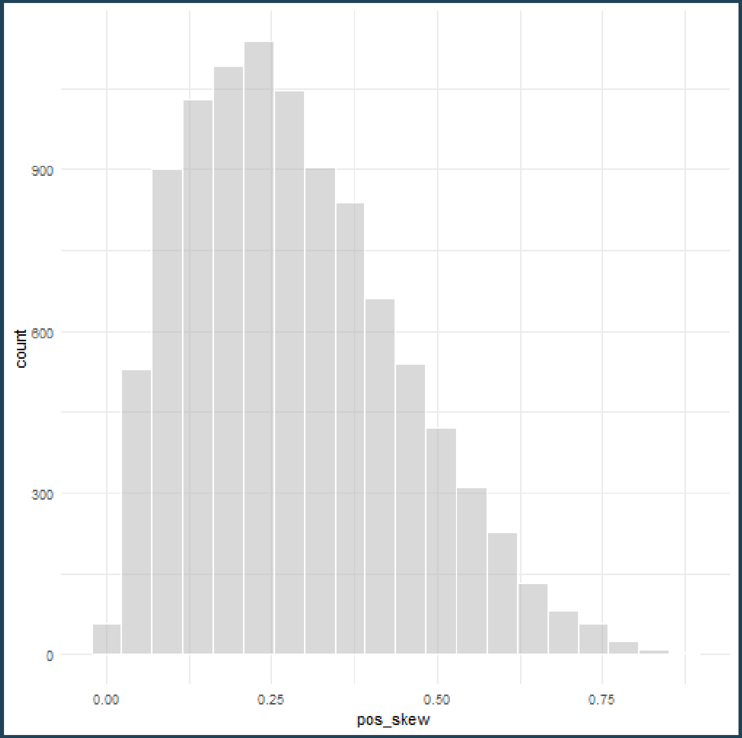
```
##              mean   sd median
## normal     -0.01 1.01  -0.03
## pos_skew    0.29 0.16   0.26
## neg_skew    0.71 0.16   0.74
```

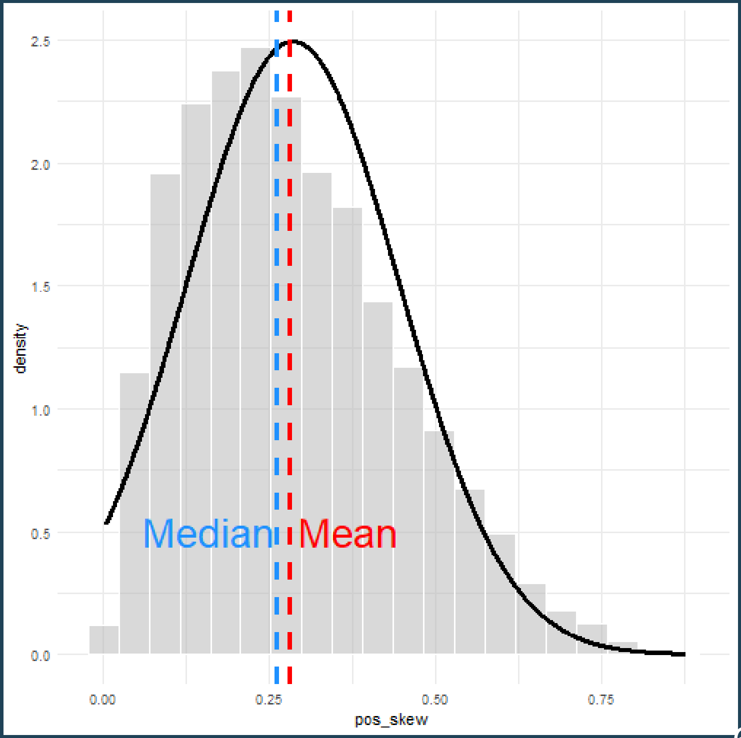




Positively Skewed Data

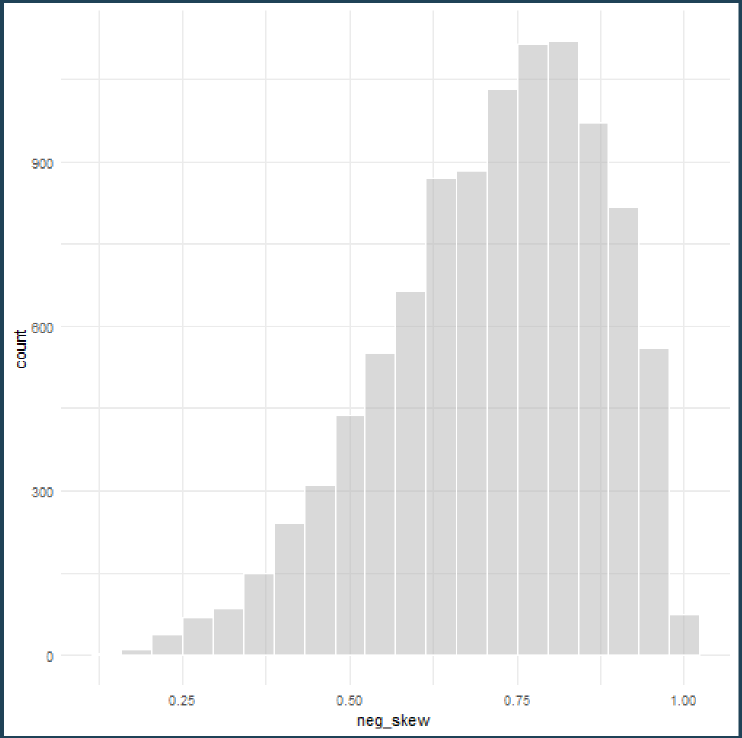
```
##             mean   sd median
## normal     -0.01 1.01  -0.03
## pos_skew    0.29 0.16   0.26
## neg_skew    0.71 0.16   0.74
```

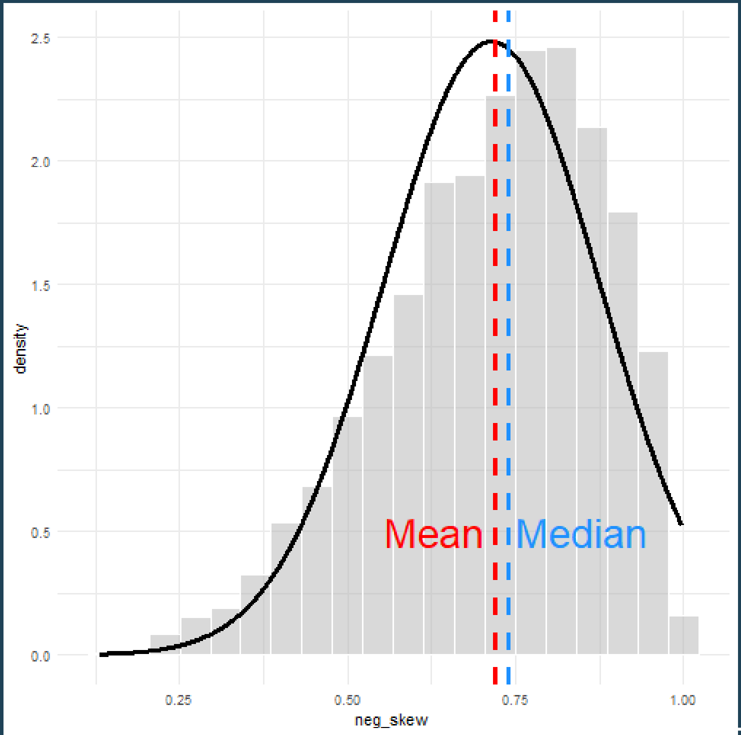




Negatively Skewed Data

```
##             mean   sd median
## normal    -0.01 1.01  -0.03
## pos_skew   0.29 0.16   0.26
## neg_skew   0.71 0.16   0.74
```





The Mode

- Value/score with the highest frequency
- Essentially useless in statistics
- **Unimodal distributions**
  distribution with only one hump; one value is the mode
- **Bimodal distributions**
  distribution with two humps; two values with the highest frequency; two modes

```
## x
## 19 20 21 22 23 24 25 26 33 34 37
##  4  4  6  2  4  3  2  2  1  1  1
```

The Median

- **Median** is the middle value/score; the 50th percentile
- Unlike the mode, it will always be close to the middle of a distribution
- You'll only ever have one median
- The symbol is:

$$Mdn = Median$$

The Median

- Preferred for ordinal/ordered data
- Not the best option for normally distributed interval & ratio scores
- Is more reliable when dealing with skewed data
- *Important Note* If you have an even number of scores/values, then you will add the two middle values and divide by 2

```r
odd <- c(1, 6, 3, 8, 9, 8, 3)
even <- c(1, 6, 3, 4, 9, 8, 5, 6)

sort(odd)
```

```
## [1] 1 3 3 6 8 8 9
```

```r
sort(even)
```

```
## [1] 1 3 4 5 6 6 8 9
```

```r
get_median <- (5+6)/2
get_median
```

```
## [1] 5.5
```

```r
median(odd)
```

```
## [1] 6
```

```r
median(even)
```

```
## [1] 5.5
```

The Mean/Average

- **Mean** is the score located at the `mathematical` center of a distribution
- Xbar is often used for the mean
  - $$\overline{X}$$
- $$\overline{X} = \frac{\Sigma\ X}{N}$$

  is the formula to calculate the mean.
  - Xbar is the Sum of X/Scores divided by the total number of observations/scores/values
- Calculate the mean for interval and ratio scales
  - The mean of ordinal/ordered data makes no sense
- Basis for most `inferential statistics`

```r
x <- c(21, 21, 25, 23, 24, 23, 20, 19, 24, 25, 21, 21, 26, 22, 22, 24, 2

sigma_x <-  sum(x)
N <-  30

x_bar <- sigma_x/N

x_bar
```

```
## [1] 23.2
```

```r
mean(x)
```

```
## [1] 23.2
```

When to Use Each Measure

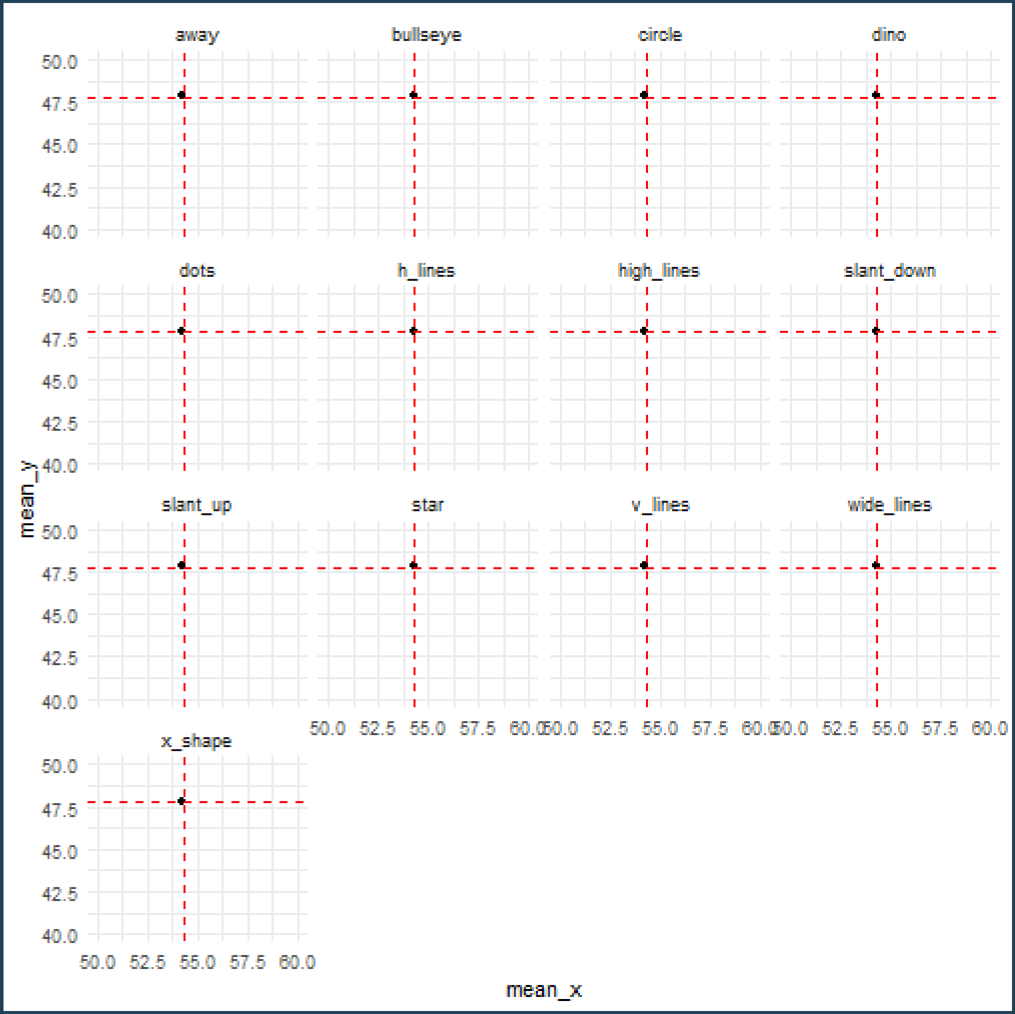
- The Median more accurately describes/summarizes skewed data compared to the mean
    - The mean will be pulled toward the extreme tail of the distribution
- Normal distribution then use the mean as the best measure to describe/summarize data

The Mean in Research

- Most statistics revolve around the mean
- Can't just trust descriptive statistics like the mean

Descriptive Statistics Are Never Enough

```
## Warning: package 'datasauRus' was built under R version 4.0.5
```


away
bullseye
circle
dino
dots
h_lines
high_lines
slant_down
slant_up
star
v_lines
wide_lines
x_shape
50.0
47.5
45.0
42.5
40.0
50.0
52.5
55.0
57.5
60.0
mean_y
mean_x

Deviation

- The distance a participant's score/value is from the mean/average
- Deviations can be positive or negative
    - participants can score lower (negative) than the mean and higher (positive) than the mean
- To get the deviation, you subtract the mean from each participant's score
- $$X - \overline{X}$$
- The larger the value the farther away from the mean the score/value is
- **Sum of the deviations around the mean** is the sum of all differences between the scores and the mean

Example

```r
x
```

```
##  [1] 21 21 25 23 24 23 20 19 24 25 21 21 26 22 22 24 20 19 19 20 20 23 21 2
## [26] 23 19 26 34 33
```

```r
one_part <- 37
another_part <- 19

x_bar
```

```
## [1] 23.2
```

```r
one_part - x_bar
```

```
## [1] 13.8
```

```r
another_part - x_bar
```

```
## [1] -4.2
```

```r
all_deviations <- x - x_bar
all_deviations
```

```
##  [1] -2.2 -2.2  1.8 -0.2  0.8 -0.2 -3.2 -4.2  0.8  1.8 -2.2 -2.2  2.8 -1.2
## [16]  0.8 -3.2 -4.2 -4.2 -3.2 -3.2 -0.2 -2.2 -2.2 13.8 -0.2 -4.2  2.8 10.8
```

```r
sum_deviations <- sum(all_deviations)
sum_deviations
```

```
## [1] 0.000000000000002131628
```

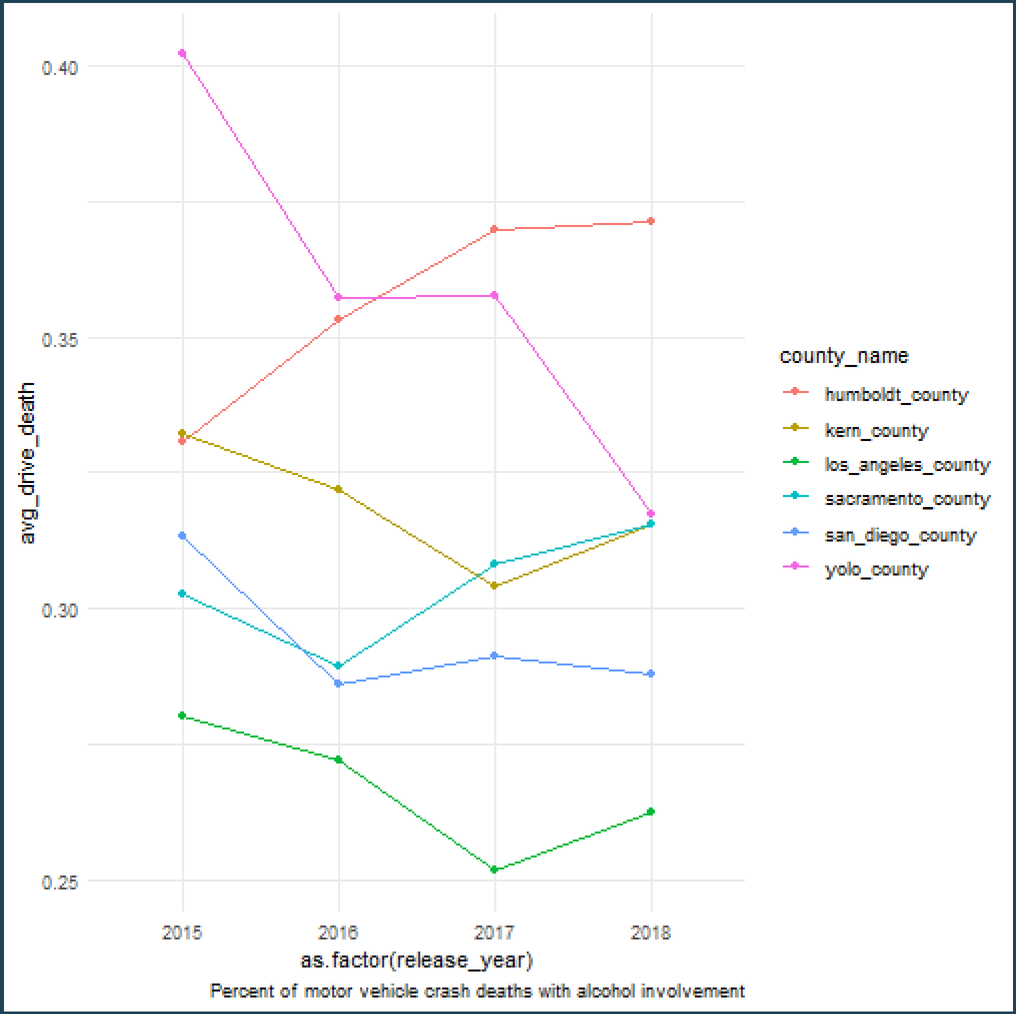
Looking to the Future

- Deviations is the start for upcoming lectures and statistical tests, especially the sum of the deviations
- $$\Sigma(X - \overline{X})$$
- If the sum of the deviations is 0 then that means your math is good
- Deviation of each score/value from the mean is often referred to as error/residual in statistical tests
- Correlational designs use the mean of IV and the mean of DV to look for a relationship between the two variables
- Experiments compare the two or more groups (IV) and the relationship with the mean value/score of the DV

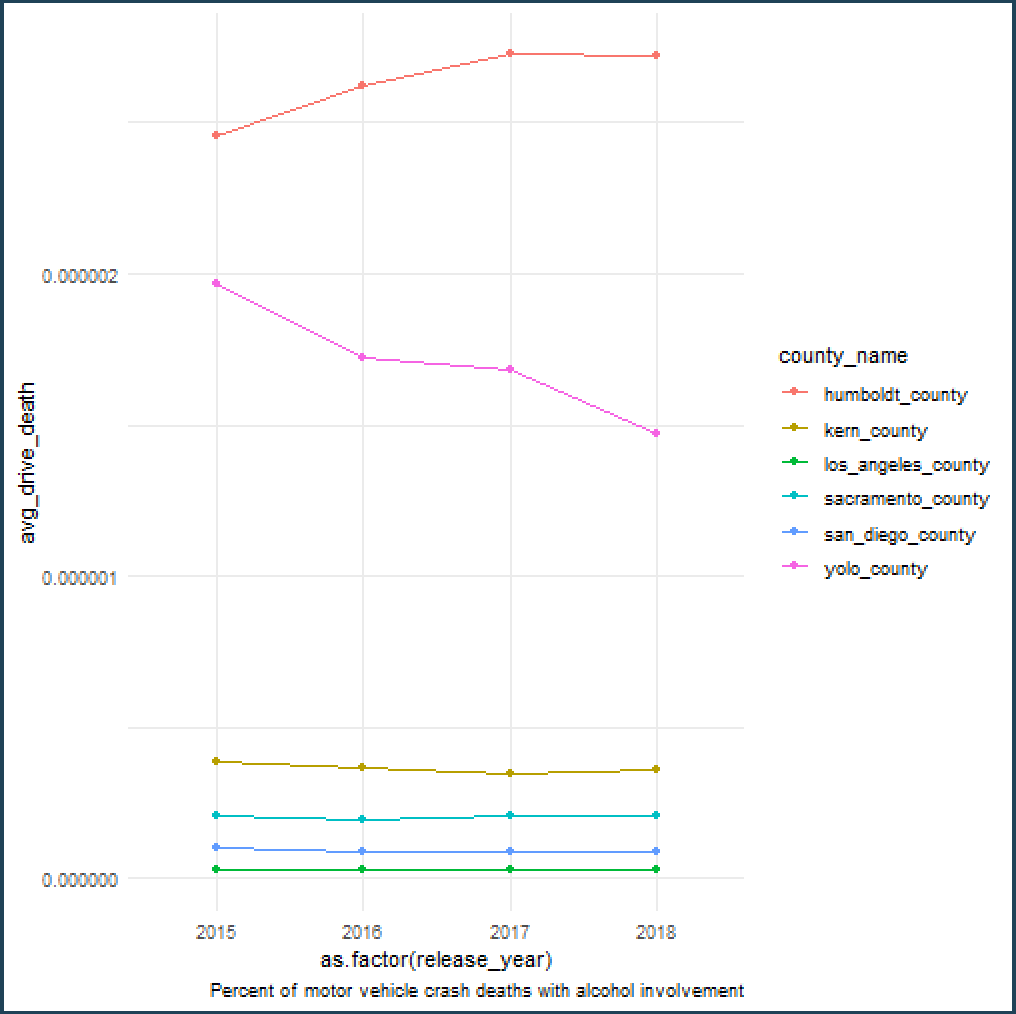
Visuals

- Line Graphs are good for showing progress or change in time
- To show group differences (nominal or ordinal IV), bar graphs are the norm
- Scatterplots are best for continuous IV and continuous DV (interval or ratio)

```
## `summarise()` has grouped output by 'release_year'. You can override using t
```



```
## `summarise()` has grouped output by 'release_year'. You can override using t
```



Describing the Population Mean

- $$\mu = Population\ Mean$$
- If you test a population then you would use mu instead of xbar
- $$\mu = \frac{\Sigma\ X}{N}$$