# *Packet 2*

*Todd CadwalladerOlsker*

*2022-01-27*

## *Descriptive Statistics*

In order to describe a data set, we need to *summarize* it. The phrase "Exploratory Data Analysis" is used in *Introduction to Modern Statistics*. We can summarize data by visualizing it, describing it numerically, or (even better) doing some of each.

I'll try to keep these packets brief, they won't contain everything there is to say about the material. You should reference the textbooks and keep your own notes as well.

### *Numeric Variables*

When we look at numeric variables, we can look at *median*-based statistics or *mean*-based statistics. The *median* is the value of the middle data point (or when there are an even number of data points, the value halfway between the two middle data points). In R, we can find the median of a data set with:

```
median(county$poverty,na.rm = TRUE)
```

```
## [1] 15.2
```

Here, the `na.rm = TRUE` tells R to ignore values that have missing data. (Try the command without the `na.rm = TRUE` and see what happens!) You can also just use `median(county$pop2017,TRUE)` for the same result.

Now, a *measure of central tendency* like the median is not worth very much without a *measure of spread* to go along with it. For the median, one way to get a handle on the spread is to also report the maximum, minimum, and 1st and 3rd quartile values. Together, these are called a *five-number summary*.

```
fivenum(county$poverty) # The pop2010 variable doesn't have any missing data, so we can skip the na.rm
```

```
## [1]  2.4 11.3 15.2 19.4 52.0
```

```
summary(county$poverty) # Is a little friendlier to the eyes, and also includes the mean.
```

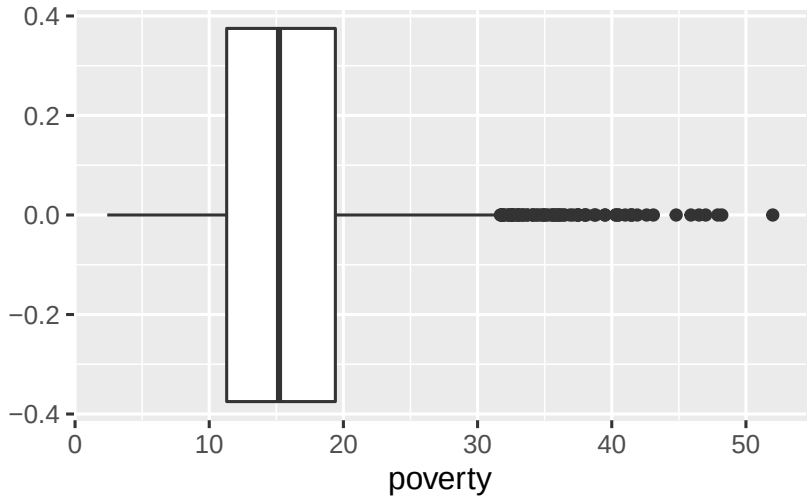
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    2.40   11.30   15.20   15.97   19.40   52.00       2
```

We can visualize the five number summary as a plot:

```
county %>% ggplot(aes(x = poverty)) +
  geom_boxplot(geom = "errorbar")
```

```
## Warning: Ignoring unknown parameters: geom
```

```
## Warning: Removed 2 rows containing non-finite values (stat_boxplot).
```




A couple of things are going on here: First, notice that there are some “dots” happening.