

## Week 5, Day 2

Sarah Moore

2022-10-20

So far...

We have established some of the following *loose* guidelines.

# Our visualizations are based on the *kind* of data we have

- 1) Continuous data (AKA quantitative data, numerical data)– to reiterate this is data that has a “natural” numerical form or measurement.

```
bigfoot$dew_point[1:45]
```

```
## [1]      NA      NA 65.72      NA      NA 51.03      NA 67.34 32.55      NA 54.
## [13] 31.01  9.50      NA  0.16 27.20 66.09 13.71 22.18      NA 44.41 12.
## [25]      NA 18.33 25.89 31.13 62.68 44.53 63.32 67.34 49.55      NA 62.
## [37]      NA 25.41      NA 28.09 22.11 21.59 39.06 33.15      NA
```

## Our visualizations are based on the *kind* of data we have

- 2) Categorical data (AKA qualitative data), we might also call this discrete data in some cases, this data does not *necessarily* have a natural, direct numerical form or measurement. The exception to this is integer counts— wherein we have a direct integer form but we can't have like half a count (e.g. number of siblings).

```
bigfoot$season[1:45]
```

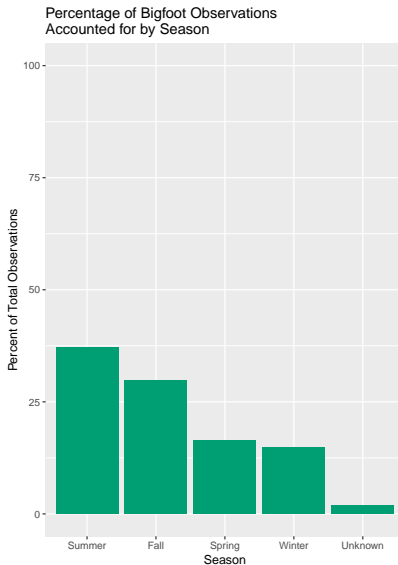
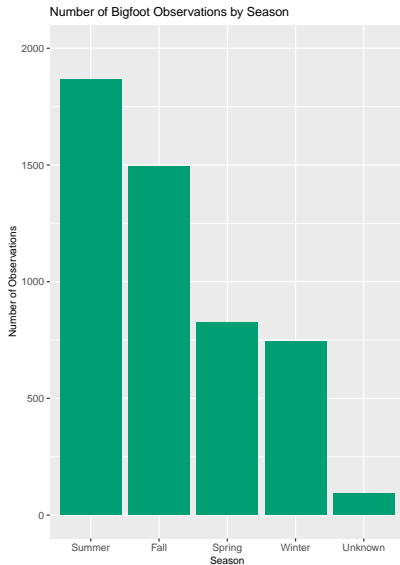
```
## [1] "Summer" "Fall"   "Fall"   "Summer" "Spring" "Fall"   "Summer"
## [9] "Fall"   "Fall"   "Summer" "Spring" "Spring" "Spring" "Summer"
## [17] "Summer" "Summer" "Summer" "Winter" "Spring" "Summer" "Spring"
## [25] "Summer" "Winter" "Summer" "Summer" "Summer" "Winter" "Summer"
## [33] "Fall"   "Summer" "Summer" "Fall"   "Winter" "Spring" "Winter"
## [41] "Winter" "Spring" "Summer" "Spring" "Fall"
```

We can further constrained by the way that we package that data

Type of Information	Suggested Visualization
Amounts	*bars*, dots, *heatmap*
Distributions	*histogram*, *density plot*, qq-plot, *boxplot*, *violin plot*, strip chart
Proportions	*bars*, *density plot*, mosaic plot, treemap, parallel set
Relations	*scatterplot*, bubblechart, *slopegraph*, contour plot, bins, *correlogram*, *line graph*

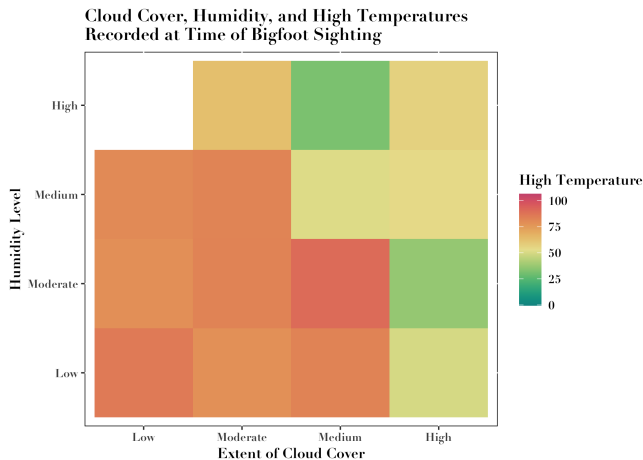
# Amounts and Proportions

## Bar Chart



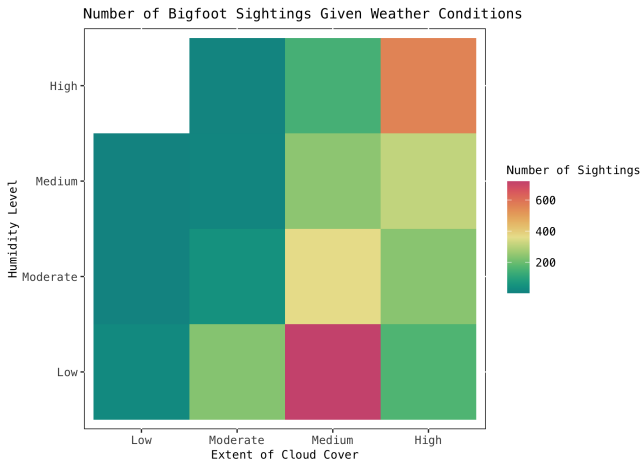
# Amounts and Proportions

## Heatmap



# Amounts and Proportions

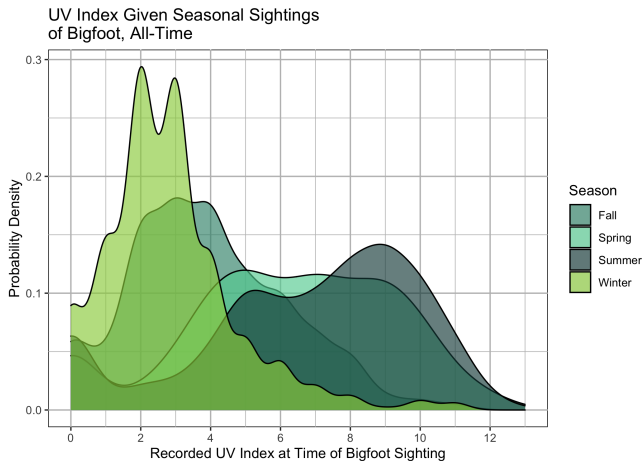
## Heatmap 2





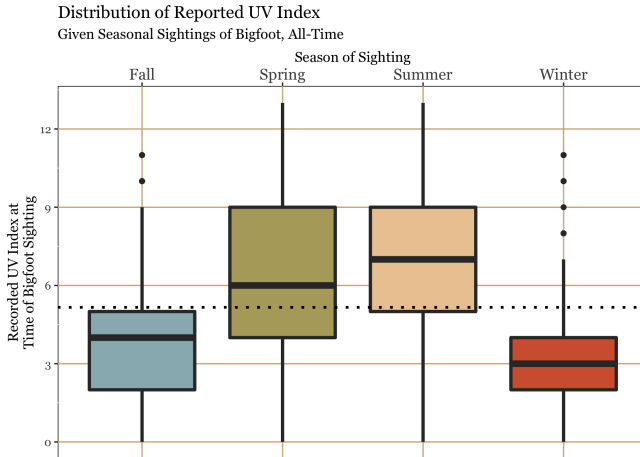
# Distributions

## Densities and boxplots



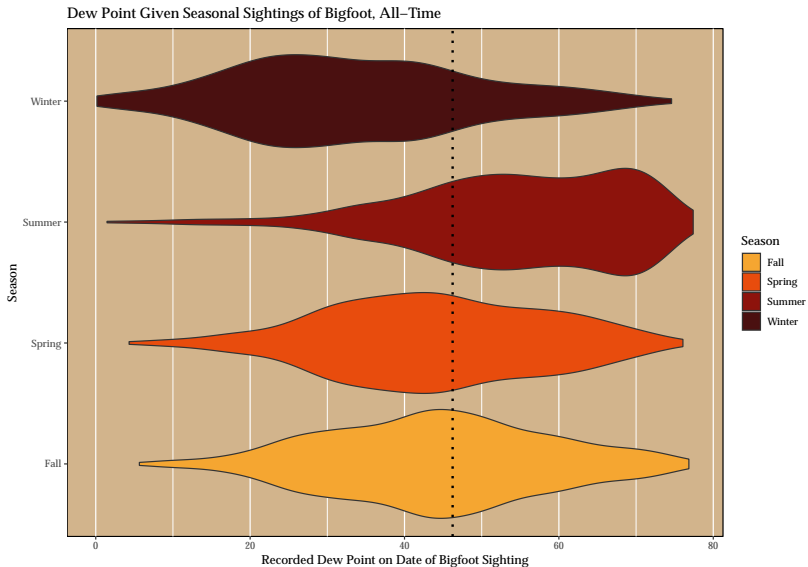
# Distributions

## Densities and boxplots



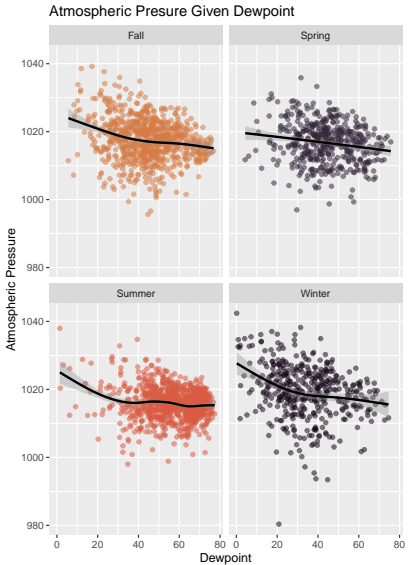
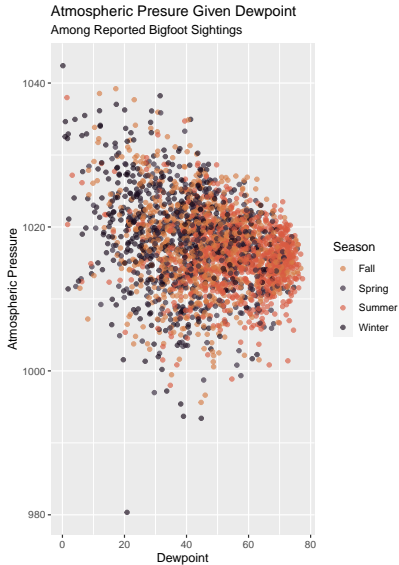
# Distributions

## Violin Plot



# Relations

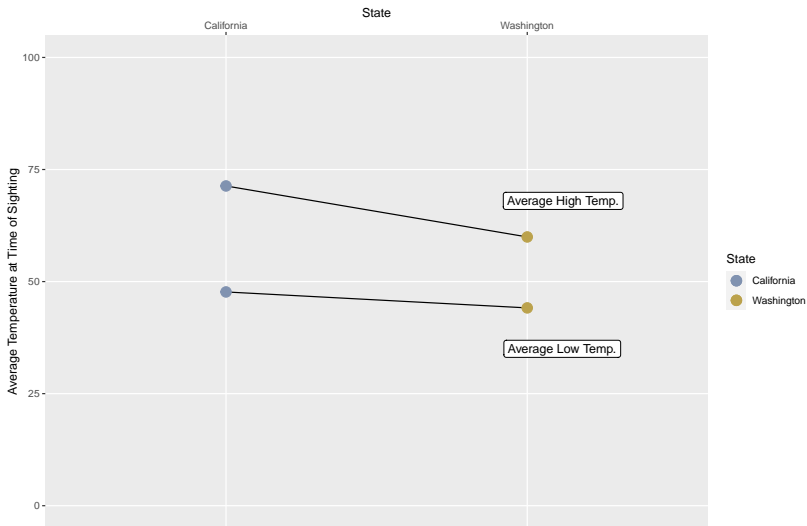
## Scatter and jitterplots



# Relations

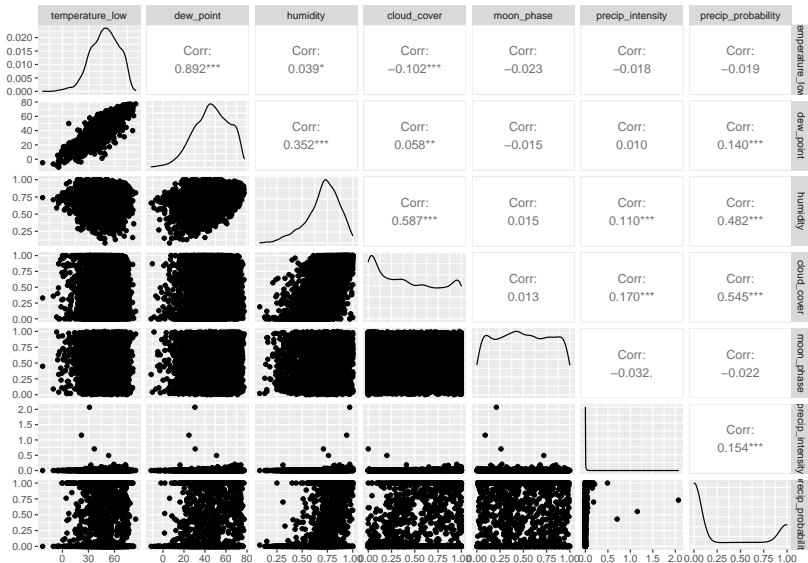
## Slopegraph

Comparison of Average High and Low Temperatures (F)  
Among States with Highest All-Time Bigfoot Sightings



## Relations

## Correlogram



And from that point we can choose to include more information.

Potential additions, usually on the basis of categorical variables OR for clarity's sake:

- ▶ Position:
  - ▶ Jitter
  - ▶ Grouping
  - ▶ Faceting/Multiples
  - ▶ Compound
- ▶ Color (either outline color or fill)
  - ▶ Discrete color
  - ▶ Gradient color
- ▶ Shape
- ▶ Transparency (either to make it easier to see density of data *or* as a function of another variable)

And we'll get to more later to account for...

- ▶ Uncertainty and statistics
- ▶ Atypical coordinate systems



## Beyond that...

We have a couple of style guidelines we have established:

- ▶ Principle of proportional ink– **the size of the ink representing data points need to be *proportional* to the data values they represent.**
  - ▶ Scales, data coverage, etc.
- ▶ Color and shapes for:
  - ▶ Distinguishing groups (categorical)
  - ▶ Represent data values (continuous)
  - ▶ Highlighting information

# Hazards

- ▶ Color and shape have limits. So, **RULE OF THUMB:**  
Qualitative color scales best for 3-5 categories, maybe less if we have some other transformations.
- ▶ Choose interpretation over ultimate aesthetic vibes.
- ▶ Provide ALT text when appropriate (i.e. a text caption in addition to your )

## Getting fancy?

- ▶ Now we have also added some more swanky elements, like high level customization with `theme()` settings.
- ▶ Also, `gganimate()`.

Anything else?

## Now what does all of this assume so far?

- ▶ A lot of what we have done so far kind of assumes that we have data that is already packaged as we want or need it. Or, I have just brushed over the code so that we can get where we want to go.
- ▶ We have worked through some ways to transform your data from one type to another (i.e. `melt()` or `pivot.wider()`), but this doesn't necessarily transform individual variables into summary points that we can plot.
- ▶ Nor do we have guidance on *what* sorts of summaries are decent, other than raw counts or proportions.