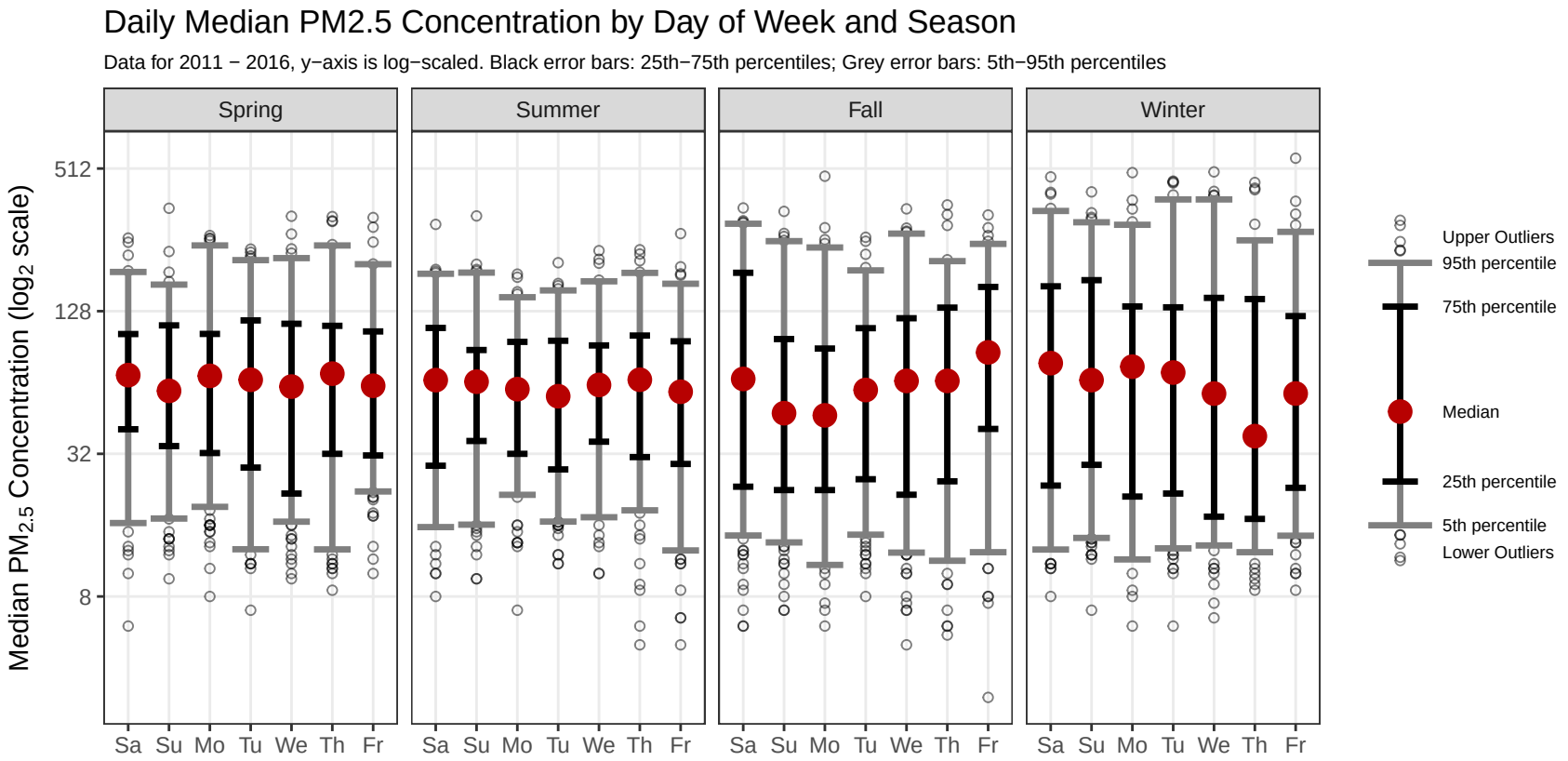
CS 765 Rough Sketches

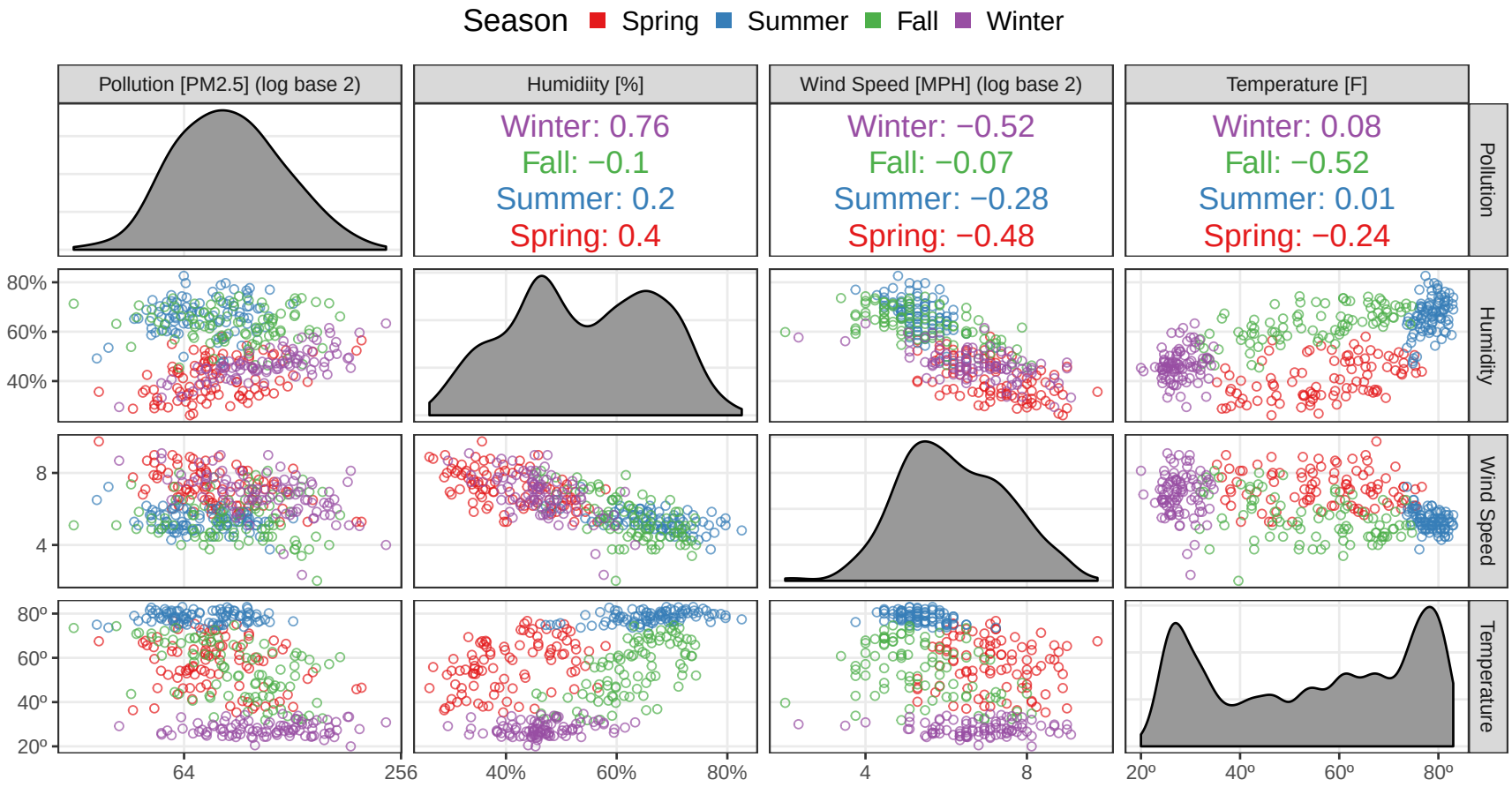
Brad Stieber



Data Daily median pollution levels (higher indicates more pollution) for 2011 - 2016 by day of week and season

Plot of daily median $PM_{2.5}$ concentration levels (measure of pollution) by day of week and season. **Red dots are the overall medians, black error bars represent 25 - 75th percentile range, and grey error bars represent 5 - 95th percentile range.** Outlier points are those that fall above or below the 5 - 95th percentile range. The y-axis is log_2 scaled.

We were originally trying to compare pollution levels by weekday / weekend. We were interested in investigating whether pollution levels were elevated during the work week. Instead, we found that although the levels remain fairly flat between seasons and day of week, the variation in pollution levels tends to increase in the Fall and Winter. Note that the error bars tend to be longer in the Fall and Winter than those in the Spring or Summer. The lengthening of error bars indicates differences in seasonal variability.

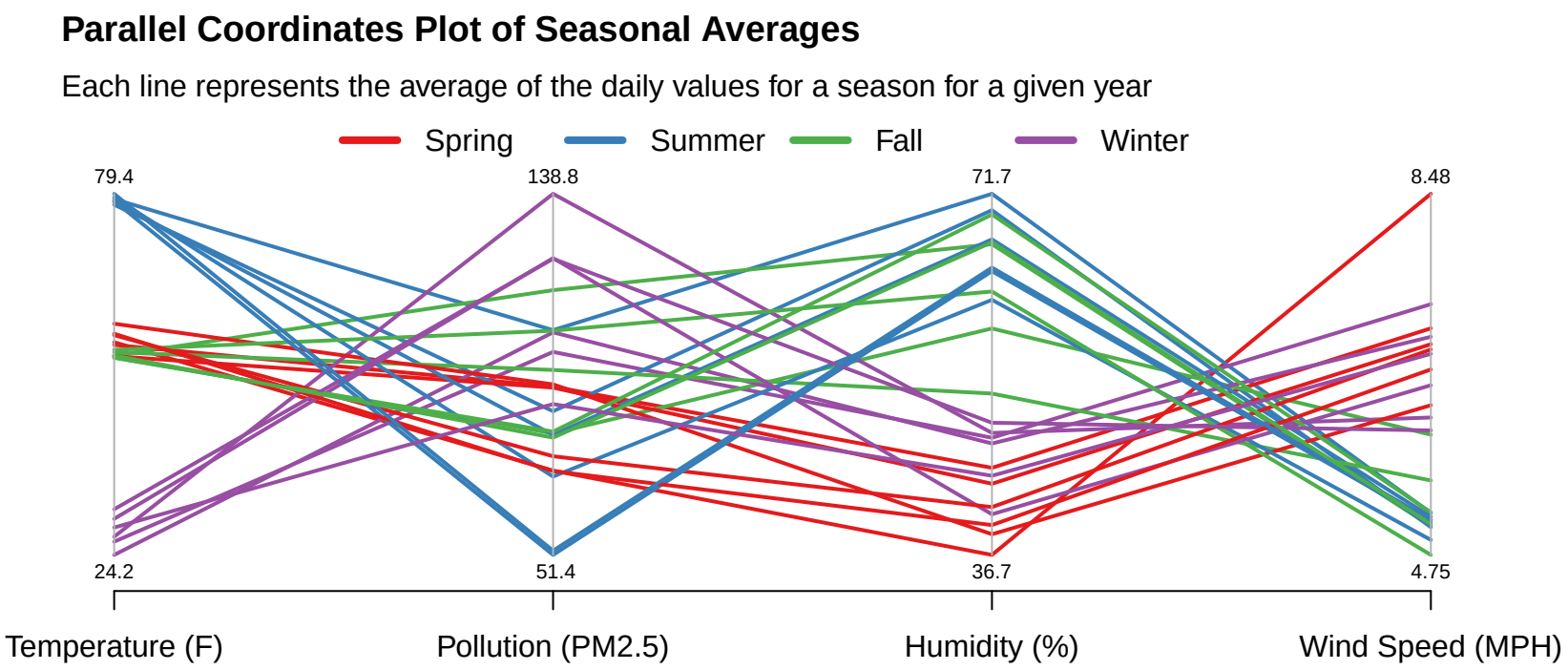


Data Average of the daily averages for pollution (higher values indicate more pollution), humidity, wind speed, and temperature for 2011 - 2016.

What relationships exist between pollution, humidity, wind speed, and temperature? Do these relationships vary depending on season? What do the distributions of these values look like? This visualization allows for many distinct queries into the complex relationships between season and our quantitative metrics.

We display a scatter plot matrix investigating pollution, humidity, wind speed, and temperature. Points are colored according to the season. Each value displayed is the average of the 2011-2016 daily averages. We choose a relatively high level of aggregation to avoid over-plotting. There should be approximately 365 marks on each scatter plot. We also display the correlations for each season between log_2 pollution and humidity, log_2 wind speed, and temperature. The axes for pollution and wind speed are log_2 scaled.

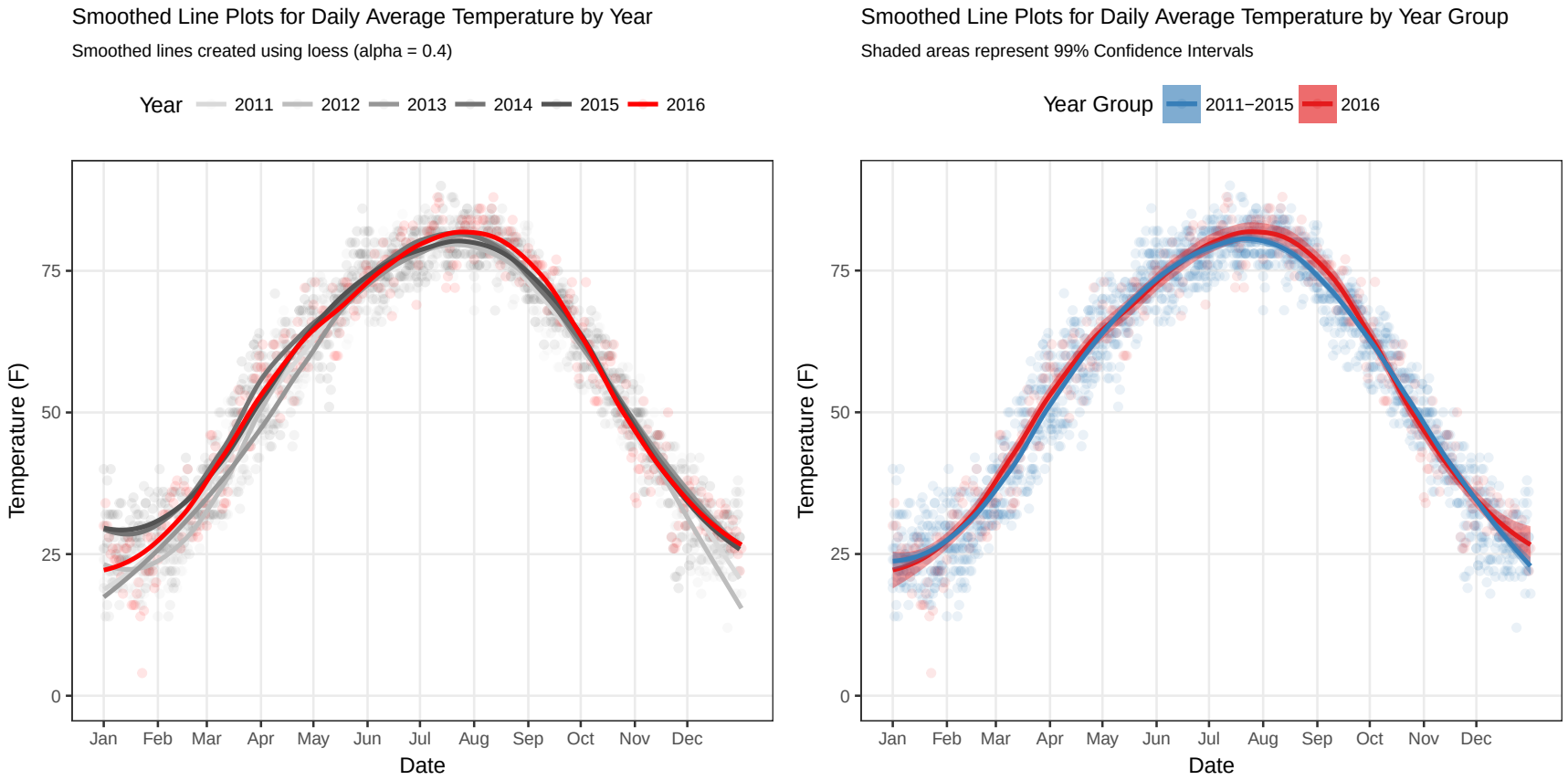
Alternative Encoding of SPLOM



Data Averages of daily averages for pollution (higher values indicate more pollution), humidity, wind speed, and temperature for 2011 - 2016, by year and season.

What relationships exist between pollution, humidity, wind speed, and temperature? Do these relationships vary depending on season? What do the distributions of these values look like?

Parallel coordinates plot investigating pollution, humidity, wind speed, and temperature. Lines are colored according to the season. Each line represents the average of the daily values for each season for each year. Though less dense than the scatter plot matrix, we are still able to ascertain some of the relationships between our quantitative metrics.



Data Daily average temperatures for 2011 - 2016 by day of year

Climate change is a major concern for many people. Even though our time frame is too short to make any clear scientific judgments, we investigate the temperature trends for 2016 versus previous years to look for signs of warming. We choose to emphasize the smoothed curves rather than the actual data to avoid spurious observations from variability in the daily data.

We display a two panel graph which utilizes a `loess` fitting method to smooth out the variability in daily average temperature data. Our left panel compares the smoothed trend for the previous years (varying shades of gray) to the smoothed trend from 2016 (red line). We color the 2016 line red to make it pop-out in the display. The right panel compares the smoothed trend for a model fit on 2011-2015 data versus a model fit on the 2016 data. The prediction from the smoothed trend for 2016 was higher than the year group trend on 241 (66%) days. In the left panel, we display the predictions with no confidence intervals, in the right panel we display the predictions as well as 99% confidence intervals.

```r
power.prop.test(p1 = 0.001, p2 = 0.0011,
                sig.level = .05, power = .95, alternative = 'one.sided')
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 2270268
##              p1 = 0.001
##              p2 = 0.0011
##       sig.level = 0.05
##           power = 0.95
##     alternative = one.sided
##
## NOTE: n is number in *each* group
```