

# M4\_assignment

Andrew Estes

2022-08-31

- 1) In addition to changes in the actual high and low temperature, take 30 minutes to explore how the variation has changed. You might look at standard deviation, IQR (remember what that is?), and you might make some box plots to explore those differences.

Below is some sample code which shows the calculation of standard deviation, variance, and IQR with self-imposed limits for one variable - max temperature. In the Module 3 Assignment I created box plots and am not including them below. I did expand upon those box plots in this assignment - you will see in the paper to local officials.

```
sd(df$TMAX, na.rm=TRUE)

## [1] 14.99752

var(df$TMAX, na.rm=TRUE)

## [1] 224.9256

lowerqTMAX <- quantile(df$TMAX, na.rm=TRUE) [2]
upperqTMAX <- quantile(df$TMAX, na.rm=TRUE) [4]
iqrTMAX <- upperqTMAX - lowerqTMAX
mild.threshold.upperTMAX <- (iqrTMAX * 1.5) + upperqTMAX
mild.threshold.lowerTMAX <- lowerqTMAX - (iqrTMAX * 1.5)
extreme.threshold.upperTMAX <- (iqrTMAX * 3) + upperqTMAX
extreme.threshold.lowerTMAX <- lowerqTMAX - (iqrTMAX * 3)

iqr <- c(extreme.threshold.upperTMAX,
          mild.threshold.upperTMAX,
          iqrTMAX,
          mild.threshold.lowerTMAX,
          extreme.threshold.lowerTMAX)
iqr

## 75% 75% 75% 25% 25%
## 158 125 22 37 4
```

## 2) Expand your technical report into either:

- a) a 1000 word White Paper, where you actually give some suggestions about what they might do differently as a result of your analysis.
- b) a 1000 word technical paper, if you don't think any change is necessary, do your best to explain the variation you've found and how it has changed over time

Imagine your audience is local officials who are using your analysis as part of a larger one about whether their local services may have to change in the future due to climate change (more snowplowing, different pool usage).

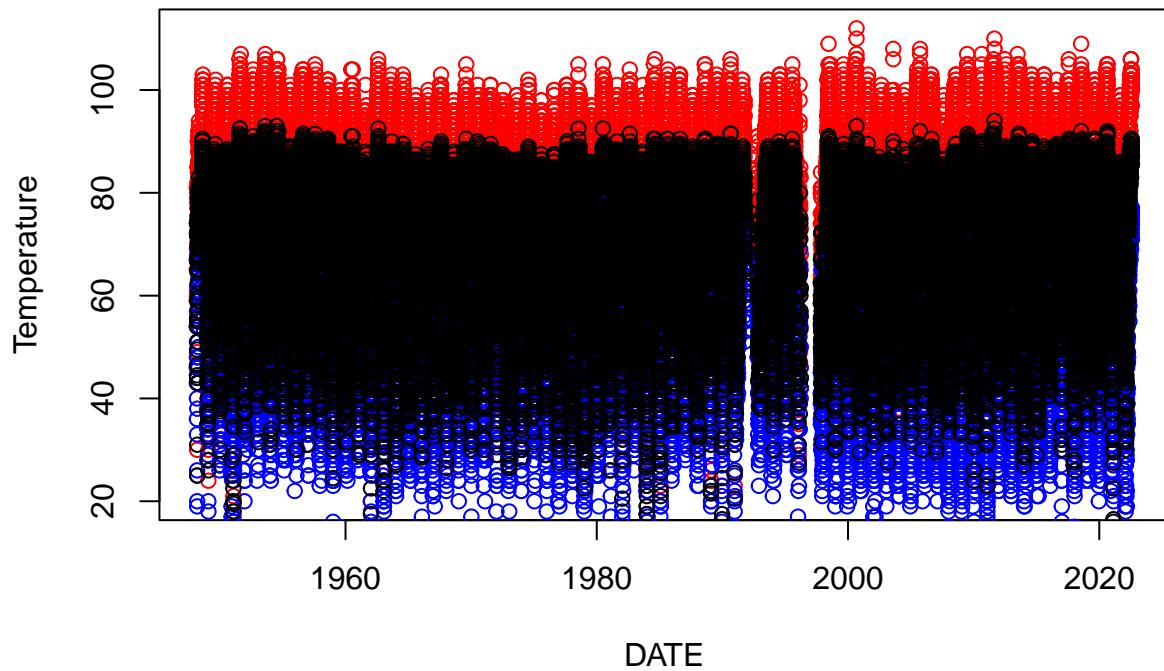
Howdy y'all,

I have been tasked with analyzing the potential climate change in the Austin, Texas area. Climate change is marked not by warming of temperatures, but by increasingly erratic weather behavior.

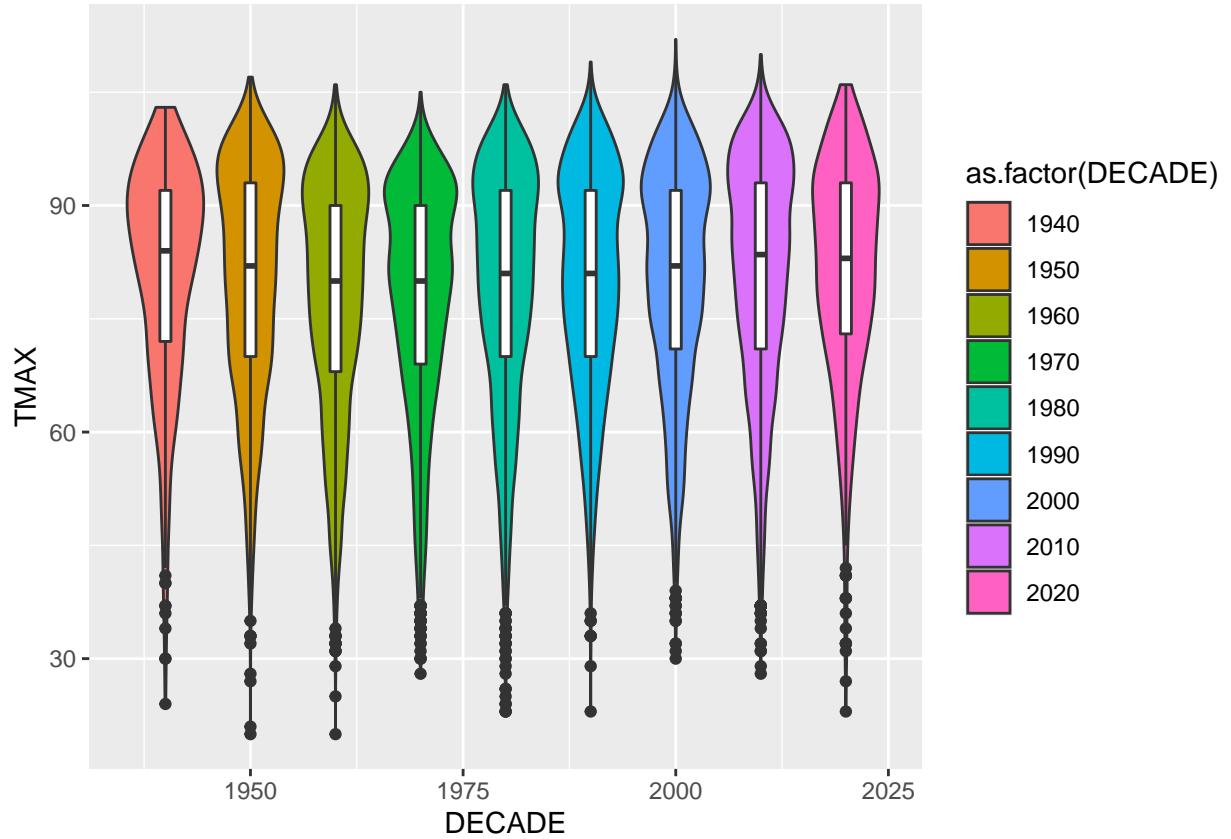
One example of climate change would be having a historically similar average minimum temperature over the course of a year, but having daily temperatures swing radically – say 30 degrees is the normal average minimum temperature with a normal range, or standard deviation, of 5 degrees. There is a statistical rule called The Empirical Rule which states that 95% of all data should fall within two standard deviations. In our example, this means 95% of all minimum temperatures should fall between 20 degrees and 40 degrees (30 degrees +/- 10 degrees). If we have 100 days of observations, that means 95 of them should fall in the 20 to 40 degree range, and 5 should be outside of it. If we have 15 days outside of it, then we can conclude that something is likely wrong and should be looked at further.

In the first chart, we are not looking at standard deviations, but we are looking at the measured daily values of maximum, minimum, and average temperature. It is pretty clear that the data from the 2000's onwards has several higher individual points than the data pre-2000.

## Scatterplot of Daily Temperature

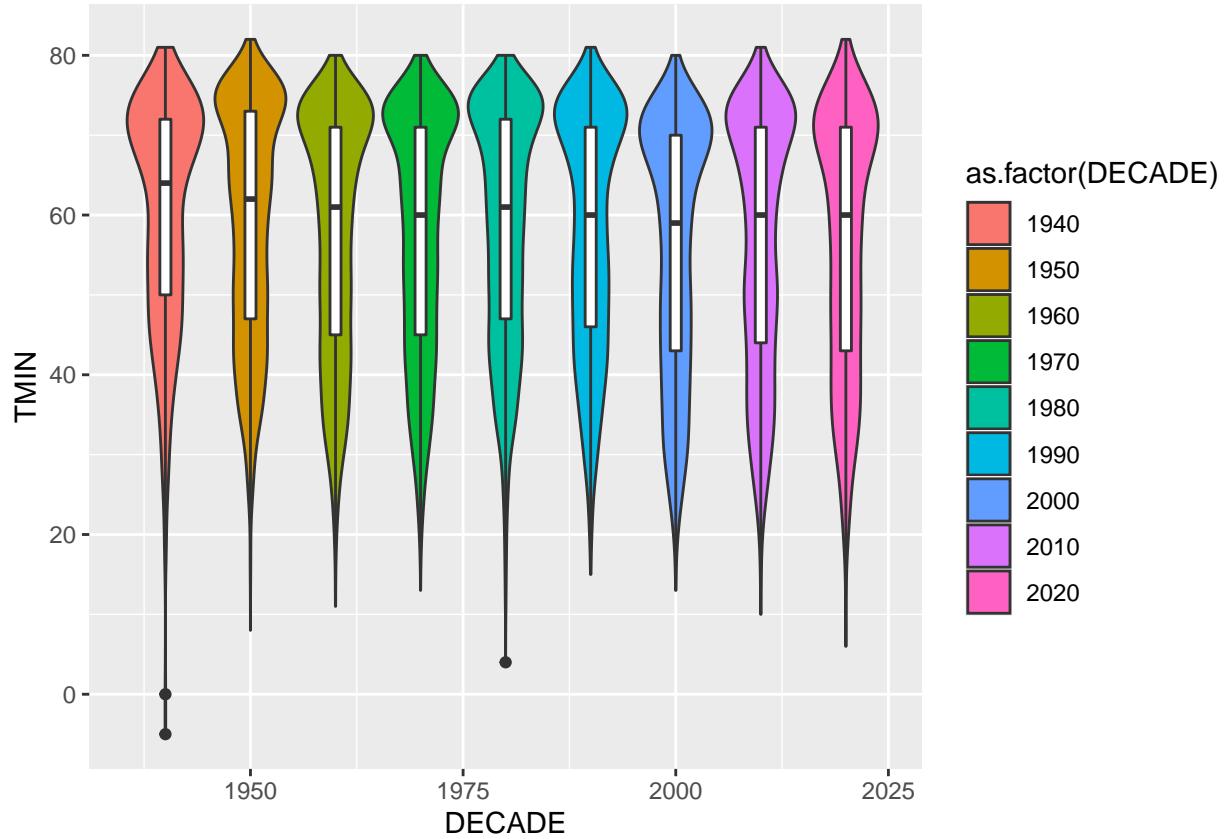


This is not indicative of any statistical significance, so we have a prettier graph to look at . . .

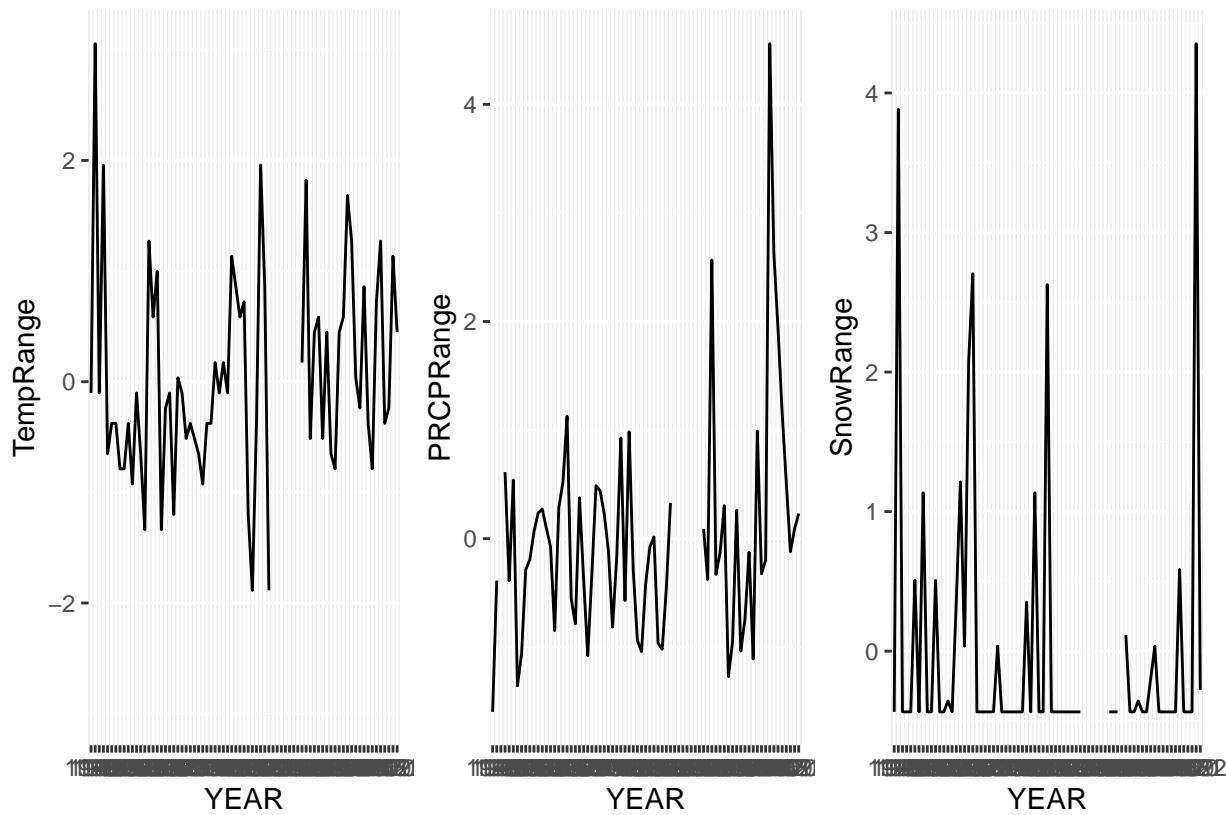


This is called a violin plot. The white box in the middle of each violin indicates the data in percentiles. The black line in the middle of the box is the median, the bottom part of the white box represents the 25% lowest value and the top part of the white box represents the 25% highest value. The shape of the violin is changed by the density of the data. Looking at the blue violin (the 2000's), it had more max temperature amounts around 95 degrees than 90 degrees which is the opposite of the 2020s (pink violin). The lines and dots extending beyond the colored violin are data points outside normal value. You can clearly see much lower than expected max temperatures in the 2020's. These outliers on the cold side are impacting the average maximum temperature and without them, we would see a much warmer average maximum temperature.

Similarly for average minimum temperature by decade, we see that the average minimum temperature is roughly the same as in years past. In the past four decades, the 1990's (light blue), 2010's (purple) and 2020's (pink) have surpassed 80 degrees as a normal minimum temperature. In the previous five decades, the 80 degree mark was surpassed in the 1940's and 1950's. A 40% surpass vs 75% surpass rate is concerning. I must agree with skeptics that this is far from conclusive – it is just something worth keeping an eye on in the future.



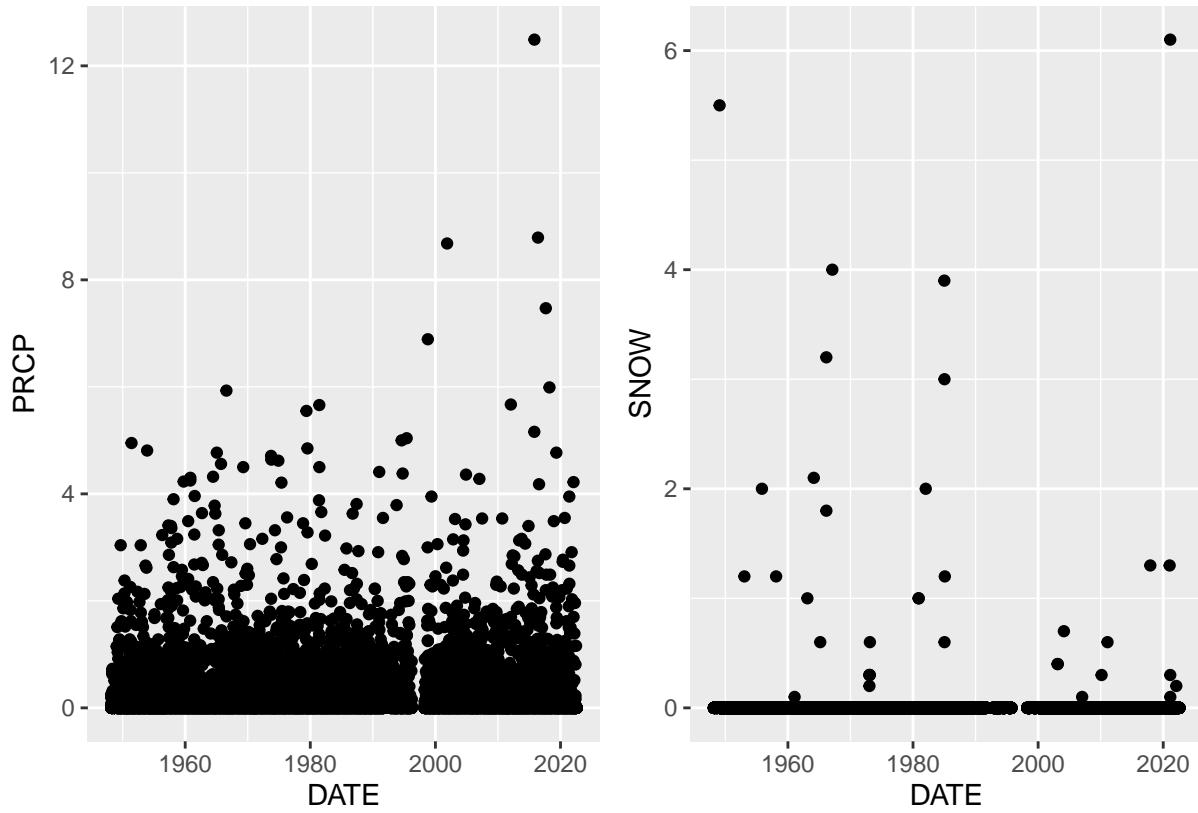
Moving from temperature to precipitation and snowfall, we run into an issue in measurability. Temperature is measured in degrees, exceeding the numerical value of 100 meanwhile precipitation is measured in inches, and does not exceed 100 – thankfully! There is a simply way to get around this issue by normalizing, or scaling, the data. In our case, we scaled the data to find center value for each variable (Daily Temperature Range, Daily Precipitation Amounts, and Total Daily Snowfall). From there, each value is turned into a standard deviation value. As you can see in the graphs below, the Range of Daily Temperatures (Maximum Daily Temp minus Minimum Daily Temperature) has actually become less varied in recent years, with all data falling within two standard deviations. However, Precipitation Amounts and Snowfall Amounts have each exceeded 4 standard deviations in recent years.



The Precipitation and Snowfall Amounts do not correlate to Temperature whatsoever. And the instances in which the Precipitation and Snowfall extended beyond two standard deviations lead to massive economic damages and, worse, fatalities.

Looking at the data's actual values, it is clear to see that higher than normal precipitation amounts are occurring more frequently since 2000. In the six decades of data pre-2000, there was not one instance of 6 inches of precipitation in a single day. In the two decades since 2000, we've had 5 instances of precipitation totaling greater than 6 inches, including a day over 12 inches.

Snowfall has undergone an opposite trend. Snowfall greater than 3 inches occurred 5 times in the 6 decades of data pre-2000. In the two decades since 2000, we've had only 1 instance of snowfall totaling greater than 3 inches. That single time, though, is an extreme outlier as it was over 6 inches of snow, greater than any snowfall amount in this dataset.



While this was a fairly rudimentary analysis, these two graphs clearly show that the record for precipitation and snowfall has occurred in the past decade, indicating more extreme weather patterns. With the increase in variability, there are several recommended steps to implement:

- 1) Greater insulation requirements for new construction to counteract the temperature extremes
- 2) Building upwards, like a less-extreme coastal town, to avoid rainwater flood damages
- 3) Building higher-density living quarters COMBINED with adequate public transportation
- 4) Incorporating more environmentally friendly features such as solar power panels
- 5) Water conservation should be prioritized as Austin utilizes aquifers
- 6) Encourage lifestyle changes such as siestas
- 7) Increase public education on safety with snow, flooding, and environmental changes

Future work should include an analysis on chaos using the Lyapunov exponent. The DChaos package seems to be an easy-to-use package for this analysis but is antiquated and not usable in the current software version. More in-depth work can include weather-impacting datasets such as population growth, income (blue collar vs white collar jobs, vehicle usage), and construction information. There may be additional weather related data as well that can be used for further research, such as local CO<sub>2</sub> levels or even data-smoothing across all the weather stations collecting data going back to the 19th century.

It could also be worthwhile to research correlated weather patterns – the weather in Kansas City today seems to become the weather in Washington D.C. around two days later. Perhaps there is a similar relationship between Austin and another area in the country. If so, perhaps something can be used at the initial location that reduces extremes at the secondary location.

### 3) In addition to changes in the actual high and low temperature and its variation, explore changes in precipitation.

(maybe you already did this)

See above.

```
newPRCP <- df %>%
  mutate(dailyTransactions = 30000*(PRCP+1))

newPRCP <- select(newPRCP, c(DATE, PRCP, dailyTransactions))
newPRCP <- na.omit(newPRCP)

leadPRCP <- newPRCP %>%
  ungroup %>%
  mutate(lead = lead(newPRCP, n=3))

a <- ggplot(leadPRCP, aes(x=DATE)) +
  geom_line(aes(y=PRCP), color="blue") +
  #geom_line(aes(y=lead$dailyTransactions), color="black") +
  scale_y_continuous(
    #sec.axis = sec_axis(~ . * 20000.420000, name = "Sales")
  )

b <- ggplot(leadPRCP, aes(x=DATE)) +
  geom_line(aes(y=lead$dailyTransactions/10000), color="black")
  #geom_line(aes(y=lead$dailyTransactions), color="black") +
  
c <- ggplot(leadPRCP, aes(x=DATE)) +
  geom_point(aes(y=PRCP), color="blue") +
  geom_point(aes(y=lead$dailyTransactions/30000), color="black") +
  scale_y_continuous(
    #sec.axis = sec_axis(~ . * 20000.420000, name = "Sales")
  )

a + b

## Warning: Removed 3 row(s) containing missing values (geom_path).
```

