



HW-2

Responses & Summary





Question 4.1

Question: Describe a situation or problem from your job, everyday life, current events, etc., for which a clustering model would be appropriate. List some (up to 5) predictors that you might use.

Answer:

1. Photographs Collection Organization:

1. Breakdown portraits pictures into Family portraits, Fashion Portraits, Art Portraits etc.
2. Classify Landscapes, Street Photography and Nature pictures

2. Search Optimization:

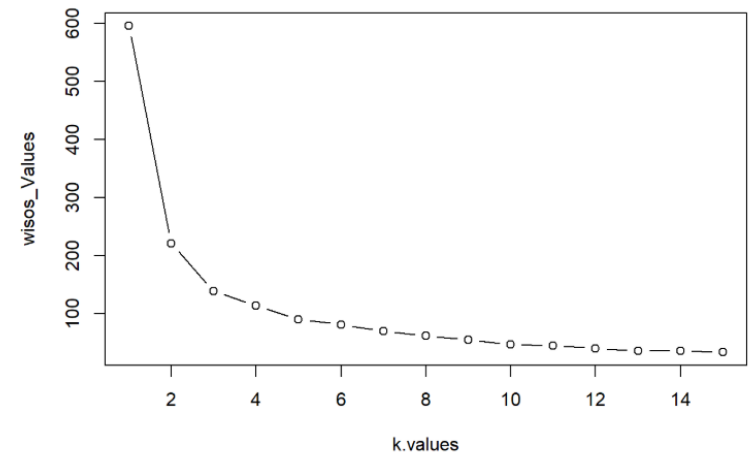
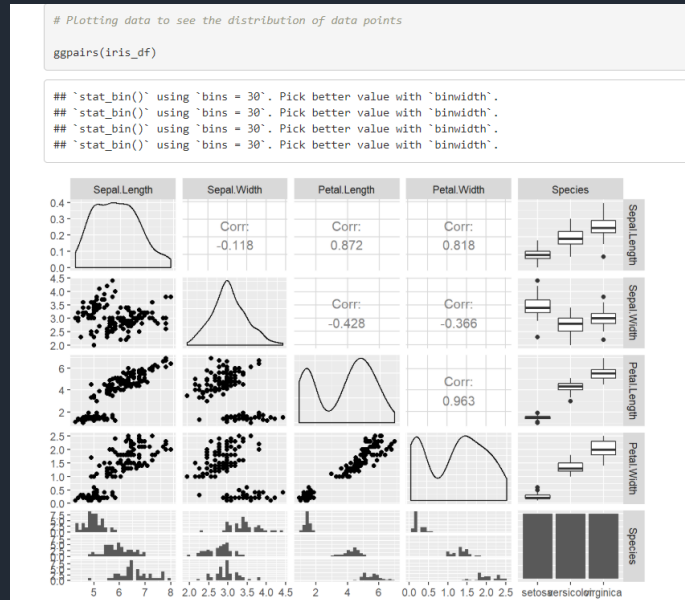
1. Recommendation of music and podcasts by time of the day
2. Recommendation of books by current topic of interest

3. Behavioral Correction: Correlate below behaviors with health data

1. Spending too much time on phone
2. Surfing or binging too much

Question 4.2

- Explored the data using ggpairs
- Analyzed for K-means 3, based on EDA observations
- Further explored for K 1 to 15
- Analyzed for optimized K-values based on SOS distance
- From this visualization it is clear that 3 clusters will suffice and there is not much significant improvement in center distance unless data is broken into too many clusters.

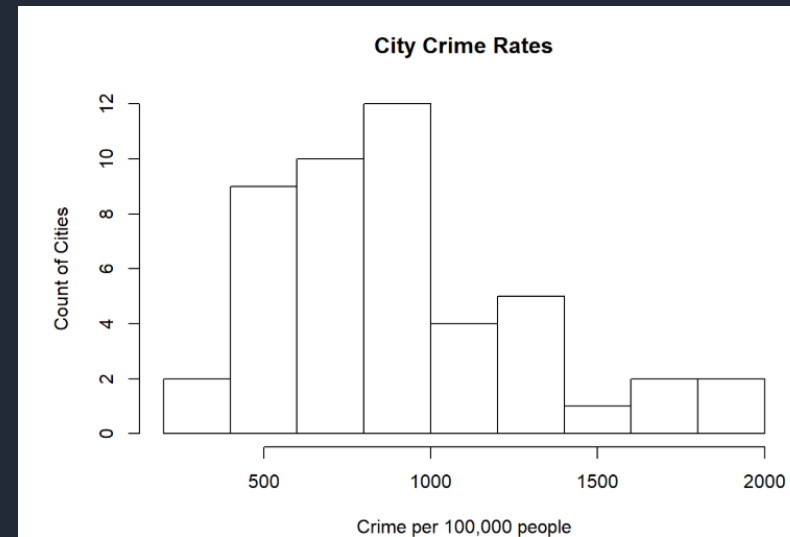


Complete Repository available on GitHub:

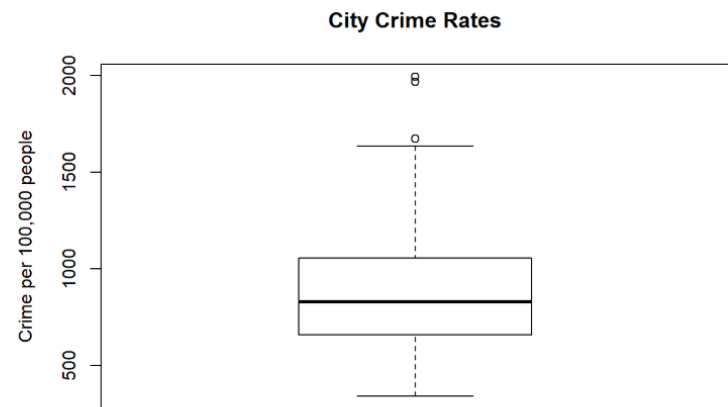
https://github.com/Hizzyth/GTX_Introduction-to-Analytics-Modelling

QUESTION 5.1 : Outlier Test

- Explored the data using histogram
- Boxplot used to visualize outlier
- Based on Outlier function, outlier was identified
- Ran Grubbs test to confirm the findings



- For Grubbs test I decided to run Type 10 as it was evident in Box plot that the outlier is on top tail only and nothing on the bottom tail end.





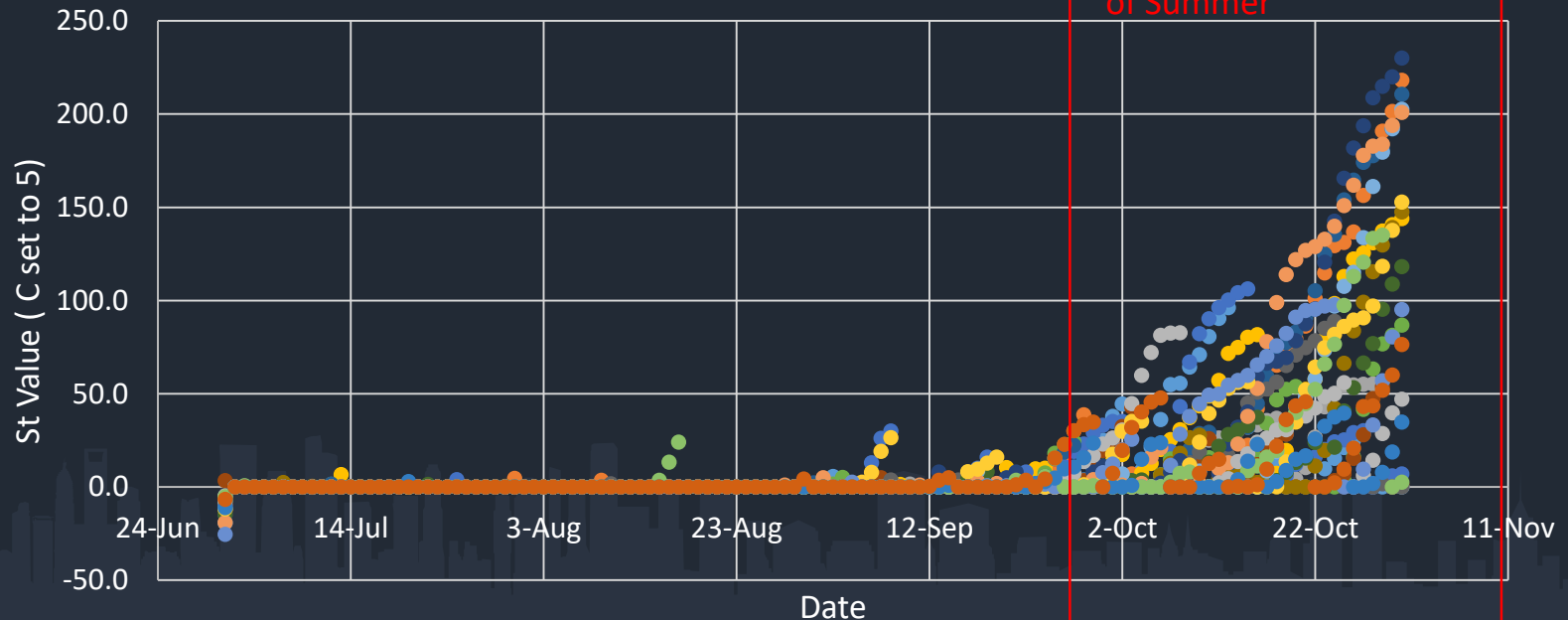
QUESTION 6.1 :: Change Detection

Question: Describe a situation or problem from your job, everyday life, current events, etc., for which a Change Detection model would be appropriate. Applying the CUSUM technique, how would you choose the critical value and the threshold?

Answer: Golf Score Improvement. Since I heavily invest my time in it, I believe that Change detection model will be perfect for pointing out any improvements or deterioration. Critical value for that will be 7 points, as that much variance is acceptable due to changing conditions (different course) and Threshold will be 12. Once the threshold is breached, time to reevaluate the game and practice sessions.

QUESTION 6.2 : Change Detection (1)

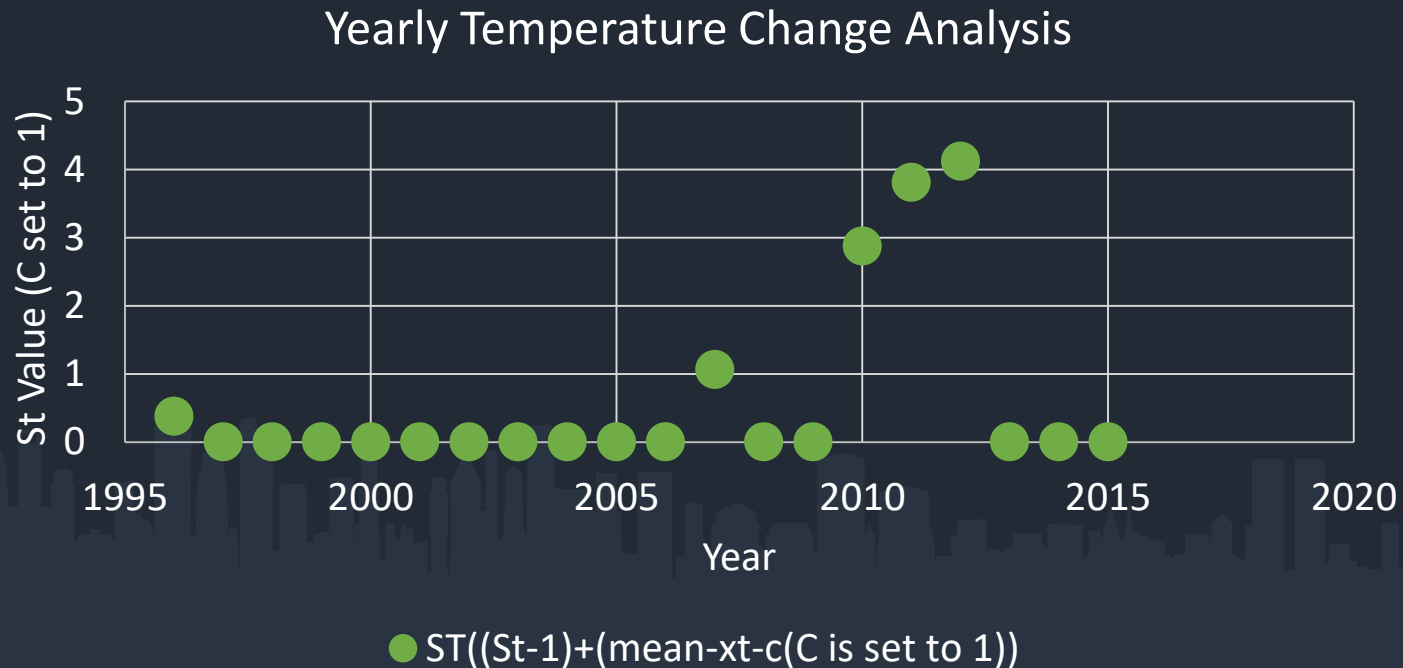
Date wise Change Analysis



● ST 1996 ● ST 1997 ● ST 1998 ● ST 1999 ● ST 2000 ● ST 2001 ● ST 2002 ● ST 2003 ● ST 2004 ● ST 2005
● ST 2006 ● ST 2007 ● ST 2008 ● ST 2009 ● ST 2010 ● ST 2011 ● ST 2012 ● ST 2013 ● ST 2014 ● ST 2015

- Used Excel for change detection
- I have taken Constant is set to 5. Further Threshold was set to 10 and from there it is evident that the unofficial summer ends towards end of September.

QUESTION 6.2 : Change Detection (2)



- For this analysis, I have taken average of each year as X_t value.
- Then Average of all years is $\mu(\text{mean})$, further $C(\text{coefficient})$ was set to 1 with Threshold(T) to 2.
- Looking at that there is evidence that the temperature is warming up and last significant change was in 2010.