Homework 1

Louise Spekking (4256778)

The data used in this analysis is data from five Kestrel 5400 sensors placed in the town of Rijsenhout, The Netherlands over the a period from June 10th to July 17th [**data**]

1. *Compute mean statistics (mean, variance and standard deviation for each of the sensors variables), what do you observe from the results?.*

   Please type your answer here.

2. *Create 1 plot that contains histograms for the 5 sensors Temperature values. Compare histograms with 5 and 50 bins, why is the number of bins important?*
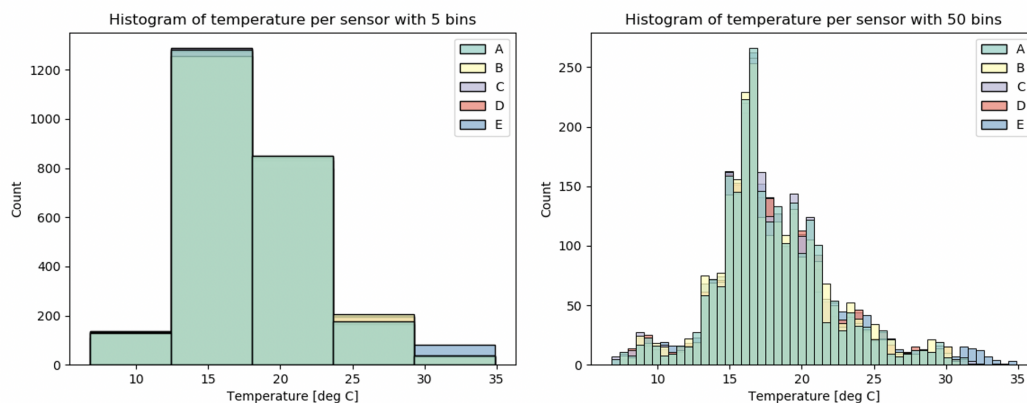

Figure 1: Histograms of temperature measurements by all 5 sensors with 5 and 50 bins.

   Please type your answer here.

   ```
   If you have code put it here
   ```

3. *Create 1 plot where frequency poligons for the 5 sensors Temperature values overlap in different colors with a legend.*

   some text and image

4. *Generate 3 plots that include the 5 sensors boxplot for: Wind Speed, Wind Direction and Temperature.*

   some boxplots

5. *Plot PMF, PDF and CDF for the 5 sensors Temperature values in independent plots (or subplots). Describe the behaviour of the distributions, are they all similar? what about their tails?*

   PMF PDF en CDF plots

6. *For the Wind Speed values, plot the pdf and the kernel density estimation. Comment the differences*

   PDF en KDE image and some text

7. *Compute the correlations between all the sensors for the variables: Temperature, Wet Bulb Globe, Crosswind Speed. Perform correlation between sensors with the same variable, not between two different variables; for example, correlate Temperature time series between sensor A and B. Use Pearson's and Spearmann's rank coefficients. Make a scatter plot with both coefficients with the 3 variables.*

   insert spearman and pearson plots

8. *What can you say about the sensors' correlations?*

   some info about the correlations

9. *If we told you that that the sensors are located as follows, hypothesize which location would you assign to each sensor and reason your hypothesis using the correlations.*

   where will the sensors be located

10. *Plot the CDF for all the sensors and for variables Temperature and Wind Speed, then compute the 95% confidence intervals for variables Temperature and Wind Speed for all the sensors and save them in a table (txt or csv form).*

    CDFs and tables

11. *Test the hypothesis: the time series for Temperature and Wind Speed are the same for sensors:*

    tabel met correlaties

12. *What could you conclude from the p-values?*

    p is low so null must go

13. *Bonus:*
    Your "employer" wants to estimate the day of maximum and minimum potential energy consumption due to air conditioning usage. To hypothesize regarding those days, you are asked to identify the hottest and coolest day of the measurement time series provided. How would you do that? Reason and program the python routine that would allow you to identify those days
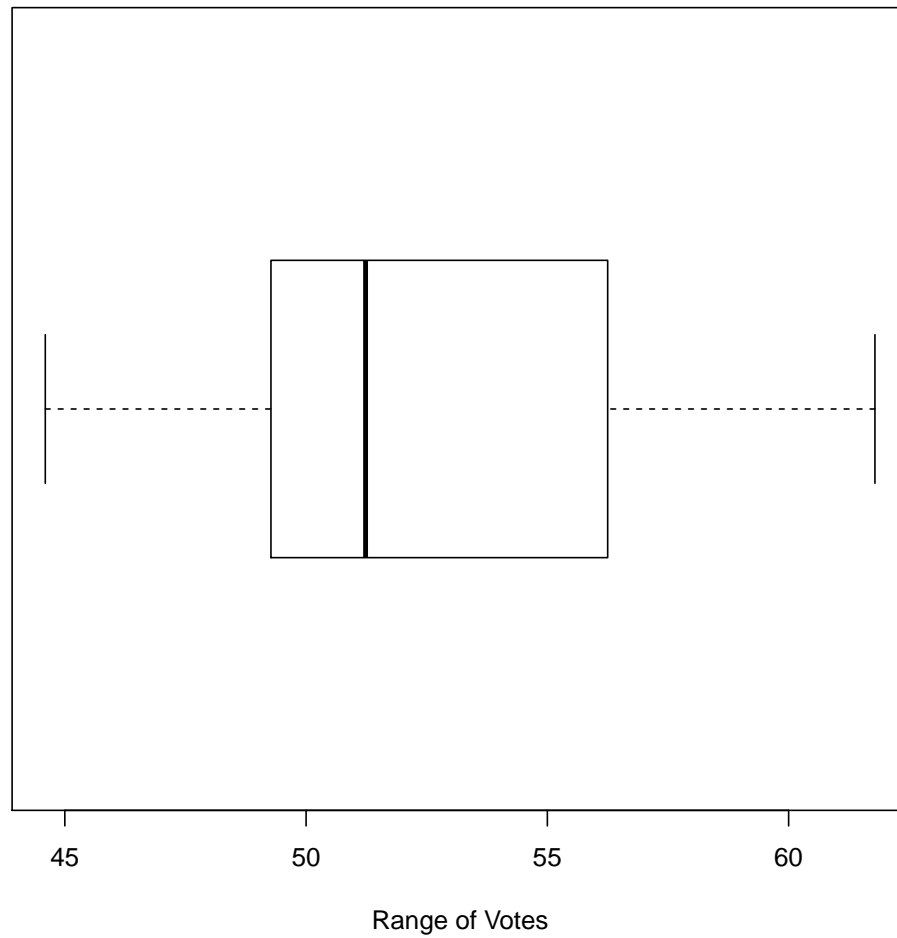
    Code van bonus question en info

A Boxplot of the Variable Vote

Figure 2: Boxplot of Incumbent Vote share

Variable	*Mean*	*Median*	*Mode*	*Var*	*SD*	*Range*	*IQR*
Vote	x	x	x	x	x	x	x
Growth	x	x	x	x	x	x	x

Table 1: Measures of central tendency and variability.