THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

STAT3622 DATA VISUALIZATION

Assignment 2, due on March 21

Use R for Question 1 and Python for Question 2.
Put all your answers and codes in a single PDF file and attach the animation .gif file.

1. For each question below, first show the plots and then interpret them.

   (a) Load the file "HKHomeCCL.csv" in R. This dataset contains home price from 1994/02/06 to 2020/3/22 for four regions including HK (Hong Kong Island), KL (Kowloon), NTE (New Territories East), and NTW (New Territories West). Show the line plot for home price versus date by regions.

   (b) Load the file "HKG_adm1.rds" in R. Show the map graph for Hong Kong. Fill the map with colors determined by "id".

   (c) Load the file "HK18Districts.csv" in R. Show the map graph for Hong Kong. Fill the map with colors determined by "Region".

   (d) Show the map graph for Hong Kong. Fill the map with colors determined by home price in 2020/03/22.

   (Hint: Use R packages, e.g., sp, reshape2. Merge datasets when needed.)

2. One data point (i.e., one number) in the *iris_bad.data* file is sabotaged and not correctly recorded.

   (a) Choose a proper style of the chart (e.g., "scatter" in the *plotly* package) to illustrate the dataset and find the incorrectly recorded data point in the chart. Add an annotation around the data point in the chart. The annotation should include all the information of the data point.

   (b) Define a function "animation(i)" and create an animation by calling the *FuncAnimation* function in the *matplotlib* package. The goal of the animation is to draw viewers' attention to the incorrectly recorded data point. For example, one can create the

animation by changing the color of the data point multiple times. Save the animation in the ".gif" format.

(c) Delete the abnormal data point and rename the data frame indexes to 1–149 (originally the index is 0–149). Print the first eight rows of this new data frame.