This problem set consists of 3 pages. Try to answer all questions.

All questions should have an answer in plain text – in some cases possibly very short. For many questions it is also relevant to show the R code you have used. Be sure to make the distinction between plain text and code clear, e.g. by using different fonts.

I. Mobility under Covid-19 (weight 50%)

Each sub-question carries the same weight.

When people use Google's positioning system, the company can track how they are moving around. To inform about response to the COVID-19 pandemic, Google has released daily data on movement patterns on categories of location. In this question, we are going to use data on time spent at home relative to a reference value calculated as the median in January and early February 2020.

The data are available as an attachment. Look at the right side.

a) Read the data into a R data frame or tibble called `mobility`. How many data points are present in the data set? The variable `date` records the date. What is the time period covered by the data set?

b) The county (*fylke*) and municipality (*kommune*) of each observation are given in the variables `sub_region_1` and `sub_region_2`.[1] These are encoded using UTF-8. Explain what that means and why this is useful to encode the names of Norwegian counties and municipalities.

c) We are going to focus on the time spent at home, given by the variable called `residential_percent_change_from_baseline`. Extract the data for Oslo from the tibble `mobility` in a separate tibble called `oslo` (counties are given by `sub_region_1`), and plot the value of `residential_percent_change_from_baseline` against time. Comment briefly on the plot.

d) Add a variable `weekday`, indicating the day of the week (Monday, Tuesday, …), as well as a binary variable for weekend days, to the tibble `oslo`. Plot the evolution of time spent at home separately for each day of the week. Next plot the values for weekend vs. weekdays over time.

e) Plot the evolution of time spent at home county by county, preferably in a common graph. Separate between weekend and weekday data to enhance readability. Identify the three spikes in the data that occur in all counties.
Hint: Aggregate county data are shown with `sub_region_2` having missing values, but with `sub_region_1` containing the name of the county.

f) Compute the mean and median response by region over the period March 12 to October 31. Which region had the largest and smallest deviations from the "normal" (i.e. the value 0, the median value observed in January)?

g) Construct a tibble with `residential_percent_change_from_baseline` for Oslo divided by the value at the national level. Show its evolution over time.
Hint: Data at the national level are shown in the data with `sub_region_1` missing values.

---

[1] Norway currently has 11 counties, Agder, Innlandet, Møre og Romsdal, Nordland, Oslo, Rogaland, Troms og Finnmark fylke, Trøndelag, Vestfold og Telemark, Vestland, and Viken.

II. Utility maximization and consumer demand (weight 30 %)

Each sub-question carries the same weight.

A consumer consumes two goods, x and y. She has income R and faces prices p_x and p_y. Her utility function is given by $U(x,y) = \dfrac{x^{1-\gamma}}{1-\gamma} + \psi \dfrac{y^{1-\theta}}{1-\theta}$. Assume $\gamma = 0.5$, $\theta = 0.3$, and $\psi = 2$.

   a) Construct a function that computes the agent's utility for a vector of values (x,y). Find the utility of $(x,y) = (2,2)$.
   b) Let $= 10, p_x = 1$, and $p_y = 2$. Plot the consumer's utility for varying levels of $x \in (0.1; 5)$, letting y be given by the budget constraint. Find the consumer's demand for the two goods.
   c) Maintain $p_x = 1$, and $p_y = 2$. Draw the Engel curve for good x, i.e. the relationship between demand and income R.
   d) Let $R = 10$ and $p_y = 2$. Draw the demand curve for good x, i.e. the relationship between demand and the price.

III. Communicating data (weight 20 %)

Each sub-question carries the same weight.

Below are two graphs. For each of the two graphs, discuss to what extent they are good at communicating the data at hand. Make sure to explain whether they can give misleading impressions of the underlying data. For each of the two graphs, make an improved graph to show the data at hand as clearly as possible.
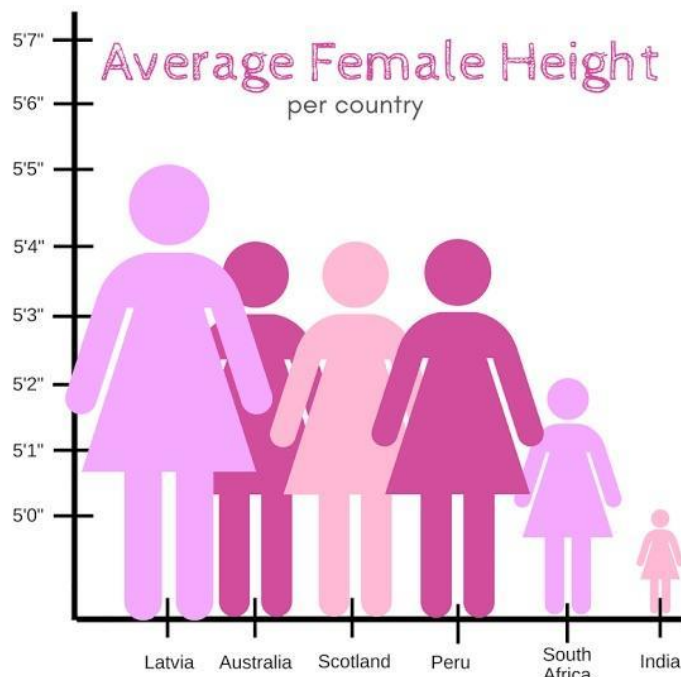*You can make this graph in R or any other statistics or design software you prefer or by sketching by hand. You can take the actual numbers shown as correct.*

a) Representing speed scores
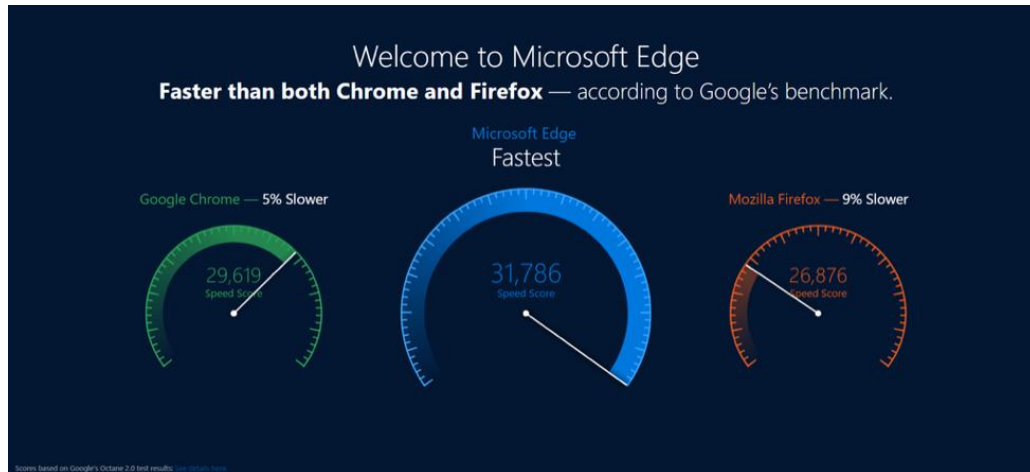
Source: Microsoft Edge opening screen

b) Representing average female height

Source: https://twitter.com/reina_sabah/status/1291509085855260672