February 23rd, 2024
Exam time: 2pm-3pm

1. (10 points) Define what is meant by Factors and Levels in a dataframe. In one or two sentences, describe each line in R-code below and explain the output:

```
> library(tidyverse)
> library(gapminder)
> h_countries <- c("Egypt", "Haiti", "Romania", "Thailand", "Venezuela")
> h_gap <- gapminder %>%  filter(country %in% h_countries)
> nlevels(h_gap$country)

[1] 142

> h_gap_dropped <- h_gap %>%droplevels()
> nlevels(h_gap_dropped$country)

[1] 5
```

2. (16 points) In one or two sentences, describe each line in R-code below:

a)

```
> U = runif(1)
> X = tan(pi*(U-0.5))
```

b)

```
> K = 1;
> p = 0.2;
> while(runif(1) > p)
+ K = K + 1;
> K
```

Find the distribution (i.e. Range and p.m.f. or p.d.f.) of output variable X and K.

3. (9 points) SantaBanta collects $X_1, X_2, X_3, \ldots, X_n$ of i.i.d measurements of electron-parameter. They assume that the observations follow a Exponential(θ). Find the maximum likelihood estimate for θ.

4. (5 points) Fill in the blanks (I),(II),(III),(IV), (V), so that the below code produces the given plot

```
library(tidyverse)
mpgcs = subset(mpg, mpg$class=="compact")
ggplot(data = mpgcs, mapping = aes(x = --(I)--, y = --(II)--)) +
  --(III)-- +  --(IV)--( size=0.4, alpha=0.9)+  --(V)--()
```