

Tutorial 7

Waseem

11/4/2021

```
knitr::opts_chunk$set(echo = TRUE)
```

1. Compute the Happiness Score for each individual.

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.4      v dplyr  1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.0.1      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
library(readxl)
```

```
library(survey)
```

```
## Loading required package: grid
```

```
## Loading required package: Matrix
```

```
##
```

```
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
```

```
##
```

```
##     expand, pack, unpack
```

```
## Loading required package: survival
```

```
##
```

```
## Attaching package: 'survey'
```

```
## The following object is masked from 'package:graphics':
```

```
##
```

```
##     dotchart
```

```
library(dplyr)
```

```
library(ggplot2)
```

```
happiness<- read_excel("data/happ_svy_data.xlsx")%>%as_tibble()
```

```
happiness_score_R<-happiness%>%
```

```
  select(c(1,5,6,10,13,14,19,23,24,25,27,28,29))%>%
```

```
  mutate(across(starts_with("Q1") , ~recode(., "Strongly disagree"=6L, "Moderately disagree"=5L, "Slightly
```

```
happiness_score<-happiness%>%
```

```
  select(!c(1,5,6,10,13,14,19,23,24,25,27,28,29))%>%
```

```
mutate(across(starts_with("Q1"), ~recode(., "Strongly disagree"=1L, "Moderately disagree"=2L, "Slightly
cbind(happiness_score_R)%>%
mutate(happy = rowSums(across((starts_with("Q1"))))/29)
happiness_score$ID <- seq.int(nrow(happiness_score))
head(happiness_score)%>%select(ID, happy, Q2, Q3))
```

```
##      ID      happy      Q2      Q3
## 1  1 3.241379 Singaporean/PR Off campus
## 2  2 2.827586      Foreigner On campus
## 3  3 4.103448      Foreigner On campus
## 4  4 3.310345 Singaporean/PR Off campus
## 5  5 1.551724      Foreigner On campus
## 6  6 3.793103      Foreigner On campus
```

2. Correct for non-response. Summarise the overall happiness score of our class.

Mean of happiness in class is 3.6419, meaning that on average, people are neither happy nor unhappy, according to the Oxford Questionnaire.

```
happiness_score<-happiness_score%>%
  filter(!is.na(Q2), !is.na(Q3))
students <- happiness_score %>%
  select(ID, happy, gender)%>%
  mutate(fpc = 100)
des1 <- svydesign(id=~ID, data=students, fpc = ~fpc)
pop.gender <- count(students, gender)
svymean(~happy, design=des1)
```

```
##      mean      SE
## happy 3.6419 0.0416
```

```
des1.raked <- rake(des1, sample=list(~gender), population=list(pop.gender))
svymean(~happy, design=des1.raked)
```

```
##      mean      SE
## happy 3.6419 0.0413
```

3. Present 2 plots based on the data, along with your explanations/comments.

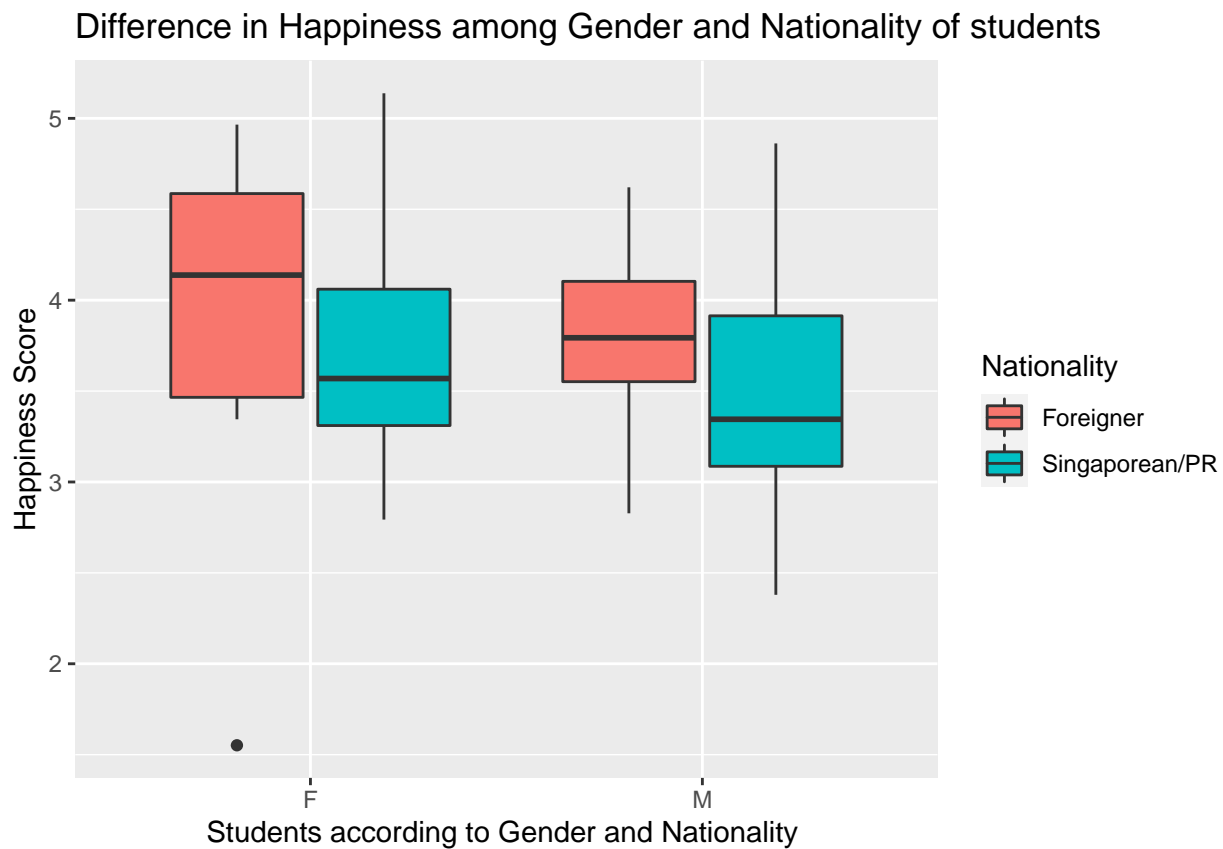
Some interesting things we could explore is how the Nationality and the gender of the students affects their happiness levels. The first plot depicts a boxplot of the happiness of foreigners and locals and their gender. Several hypotheses could be drawn from this plot.

Firstly, females are generally happier than males, as seen from the female boxplots being higher than the males. This could be due to females being more emotionally expressive, leading to them handling their emotions better than males. More information needs to be obtained to form any solid conclusion, like how often students express their emotions and how they deal with their emotions.

Secondly, among both males and females, foreigners are happier than locals. This could be due to them experiencing and enjoying things new to them, while locals are experiencing the same old culture and environment as they have been their whole life. Another interesting observation is that, the spread of happiness from the mean is a lot smaller for male foreigners compared to female foreigners. Data on what things make them happy could be used to illustrate the differences between foreigners and locals and perhaps help us be more appreciative of our own culture.

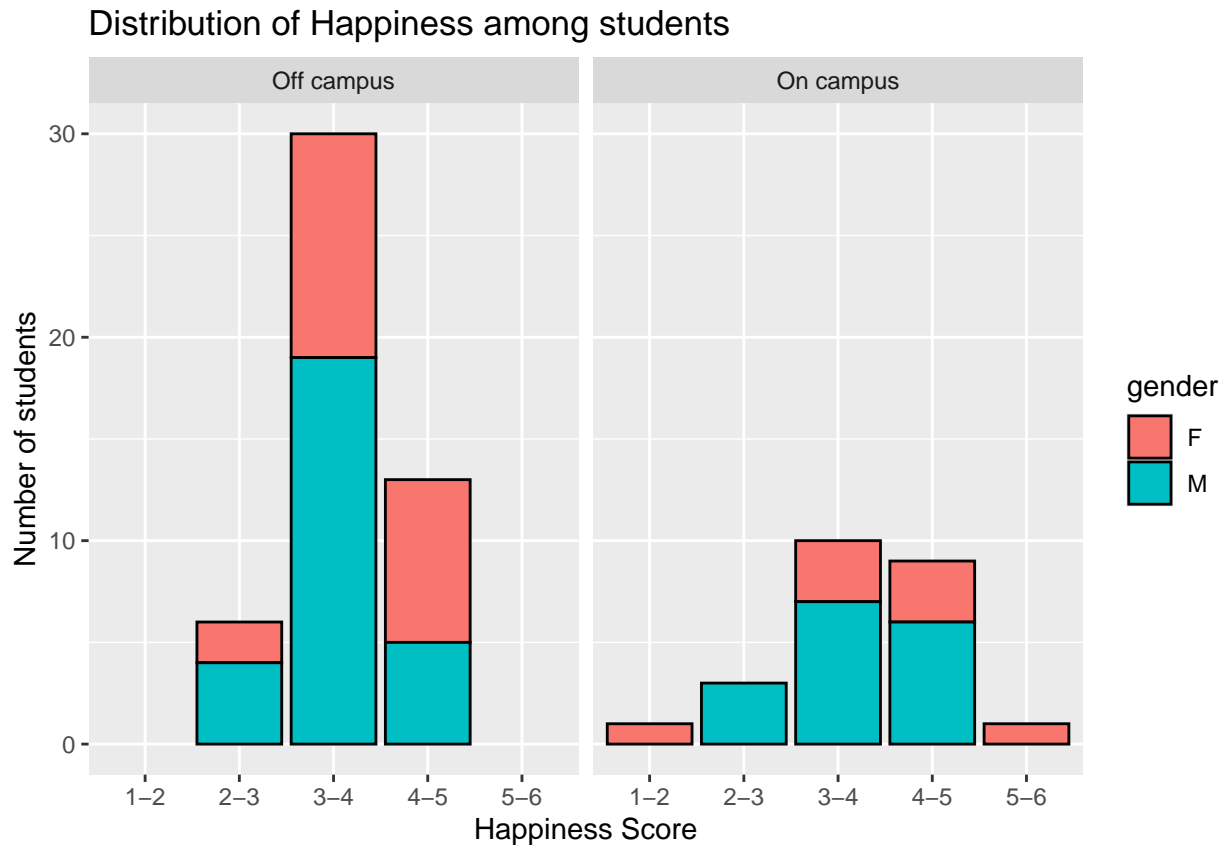
```
happyplot<-cbind(students, happiness_score$Q2)%>%
  cbind(happiness_score$Q3)%>%
  rename(Nationality= `happiness_score$Q2` )%>%
  rename(Residence=`happiness_score$Q3`)
```

```
ggplot(happyplot)+
  stat_boxplot(aes(y=happy,x=gender,fill=Nationality))+
  labs(title="Difference in Happiness among Gender and Nationality of students ", y="Happiness Score",
```



The second plot depicts the spread of happiness levels amongst the students, with fill showing the number of residents who stay on campus. Referring to the oxford questionnaire, most people are not particularly happy or unhappy; the mean is 3-4. A 3rd hypothesis can be drawn from this graph; the spread of happiness from the mean of 3-4 is wider for students who stay on campus compared to those who stay off campus. Perhaps this could be due to that people who stay on campus are exposed to more social interactions and commitments and this would dictate their lifestyle to a certain extent, which can polarize their happiness or unhappiness, while those who stay off campus have more freedom over their lifestyle and therefore do not have as much of a reason to be extremely unhappy or happy. Data showing hobbies and lifestyles ie number of hours spent studying , doing misc activities, could be used to derive further conclusions.

```
round_happyplot <- happyplot%>%as_tibble()%>%
  mutate(happy=floor(happy))
round_happyplot$happy<-recode(round_happyplot$happy, '1'="1-2", "2"="2-3", "3"="3-4", "4"="4-5", "5"="5-6")
ggplot(round_happyplot,aes(x=happy,fill=gender))+
  geom_bar(position="stack", colour = "black")+
  facet_wrap(~Residence)+
  labs(title="Distribution of Happiness among students",x="Happiness Score",y="Number of students")
```



4. Analysis: • Outlier responses

Observing the first plot, we can notice that a female foreigner has significantly low happiness levels. Upon referring to the data, Respondent ID 5 has a happiness level of 1.55 and is an outlier. This female foreigner is probably unhappy and hopefully she gets the help she needs or takes the Depression Symptoms test CES-D Questionnaire to help her understand her unhappiness and feel happier.

```
outlier_responses<-happiness_score%>%
  arrange(happy)%>%
  select(happy,ID)
head(outlier_responses)
```

```
##      happy ID
## 1 1.551724  5
## 2 2.379310 23
## 3 2.517241 67
## 4 2.724138 26
## 5 2.793103 15
## 6 2.793103 44
```

- Clusters within the class There were 3 variables that could be used for hierarchical clustering , gender, nationality and residence. By plotting graphs to compare these variables, several hypotheses and conclusions could be drawn from the data above. However, there are some limitations in the plotting of these graphs. The size of locals and foreigners were different, so the results are skewed such that differences in the two groups may be more apparent in one compared to the other due to small sample size. In my opinion, happiness is very subjective and can only be quantified to a certain extent with numbers. A better representation of happiness would require more details such as what makes them happy, how do they deal with their emotions, etc and can be put into a wordcloud similar to that in tutorial 11.