HACETTEPE UNIVERSITY
FACULTY OF ECONOMICS & ADMINISTRATIVE SCIENCES
DEPARTMENT of ECONOMICS

ECO 239
Statistics I

R ASSIGNMENT
DATA ANALYSIS
1. **Employment, unemployment, and participation rates by place of birth and sex**
2. **GPAs of Surveyed Students**

Submitted to:
Shihomi Ara Aksoy

Büşra Güleryüz
21941688

ANKARA
January 2021

I pledge my honour to that, I have not given nor received any unauthorized information to / from any individual in the process of making this homework.

Büşra Güleryüz

QUESTION 1

The dataset that is going to be analyzed in this step of assignment is named "Employment, unemployment and participation rates by place of birth and sex"[1] from OECD Stat[2] ; it is merged with GDP[3] and GDP per capita[4] of related countries and some adjustments are made.

The dataset demonstrates the employment and participation rates based on gender and birthplace, but in order to make more specific and purposeful analysis, this dataset's variables are reduced.

Base-year is **2019** and only the native-born people's participation in labour force (based on sex) of related country is analyzed. Categorical variable "Countries" is shortened into 15: Australia, Canada, Czech Republic, Estonia, France, Germany, Greece, Hungary, Lithuania, Mexico, New Zealand, Poland, Slovak Republic, Slovenia and Turkey. And 5 numerical variables are: Men, Women, Total (Showing participation of men, women and total participation of country, respectively), GDP, GDP per capita.

I have merged 3 data tables from OECD Stat and World Bank Open Data. Data in columns related to participation in labour force are from OECD and GDP related columns are from The World Bank.

There are 2 questions to be analyzed in this part of assignment:

1. Are men and women participated equally in labour force in different countries?
2. Which one of the following factors is more distinctive to determine a country's prosperity: GDP or GDP per capita? And what kind of an effect the Rate of Total Participation in Labour Force have on this? Does it affect GDP per capita more directly (relatively to other countries in the data list) than it does GDP?

The Data Table in Question

Participation rate of
- Men
- Women
- Total participation

And GDP & GDP per capita for each country.

> ➢ Sorted in descendant order for % of Total Participation in Labour Force.

Country	% Of Men's Participation in Labour Force	% Of Women's Participation in Labour Force	% Of Total Participation in Labour Force	Nominal GDP (in millions)	GDP per capita
New Zealand	84.4	77.7	81	$ 206,929	$ 42,084
Germany	83.5	77.1	80.3	$ 3,845,630	$ 46,445
Australia	83.5	76.3	79.9	$ 1,392,680	$ 55,060
Canada	81.6	76.9	79.3	$ 1,736,426	$ 46,194
Estonia	81.9	75.7	78.8	$ 31,387	$ 23,723
Lithuania	79.2	77.1	78.1	$ 54,219	$ 19,601
Czech Republic	83.2	69.6	76.5	$ 246,489	$ 23,494
Slovenia	77.6	73.5	75.6	$ 53,742	$ 25,946
Slovak Republic	78.8	66.3	72.6	$ 105,422	$ 19,266
Hungary	79.9	65	72.4	$ 160,967	$ 16,731
France	75	69.7	72.3	$ 2,715,518	$ 40,493
Poland	77.6	63.4	70.5	$ 592,164	$15,692
Greece	75.7	60.1	67.9	$ 209,853	$ 19,582
Mexico	81.9	48.8	64.6	$ 1,258,287	$ 9,946
Turkey	78.2	38.9	58.6	$ 754,412	$ 9,126

◆ Numerical Summary of The Data

a.   Are men and women equally participated in labour force?

>summary(Men)
>summary(Women)

Data summary for men:

Min	1st Qu.	Median	Mean	3rd Qu.	Max
75	77.9	79.9	80.13	82.55	84.4

Data summary for women:

Min	1st Qu.	Median	Mean	3rd Qu.	Max
38.9	64.2	69.7	67.74	76.6	77.7

- As we compare both tables, they are absolutely not. On average, approximately 80% of men and 67% of women that are living in countries in this dataset are participated in labour force.

- Absolute value of the difference between 3rd quartile (1.1%) and maximum value is smaller than the absolute difference between 1st quartile and the smallest value

(25.3%). Which means that there are small numbers of countries where women's participation rate is high; and more countries have these rates for women relatively low.

- Higher standard deviation means widespread distribution from mean. Standard deviation for participation of men is **2.97** while it is **11.37** for women. This shows that rates of women's participation are strongly varied while men's rates are closer to each other in numbers. Men all over the world participate in labour while many women have difficulty to work; they are trapped into their houses to serve their family and they don't have economic independency.

- Correlation coefficient for men and women's participation rates is: 0.3636176; meaning that there is a positive relationship between these two genders' participation rates, but it is weak.

- Minimum rate for women's participation is 38.9% with Turkey and maximum rate is 77.7% with New Zealand, doubling Turkey's rate.

b.   How does total participation in labour force effect the country's GDP? Does it affect GDP per capita more directly (relatively to other countries in the data list) than it does GDP?

When the list is sorted by total participation rate in descendant order, we see that GDP itself does not seem directly affected by it. What I mean by this is: of course, a country's GDP could increase if total participation rate increases, but I am trying to draw attention to compare these countries' GDP's, whether its amount in dollars (nominal value) depends on the participation rate respectively. GDP per capita may be more responsive to the participation rate.

➤ **Deriving covariances and correlation coefficients of required variables.**

    1. <u>Covariance</u> for <mark>Total Participation Rate and GDP</mark> is =

➤ cov(GDP,Total) = 984033.9

Meaning: Covariance > 1 => These two variables are directly proportional.

    <u>Correlation Coefficient</u> for <mark>Total Participation Rate and GDP</mark> is =

➤ cor(GDP,Total) 0.1348988.

Meaning: Correlation Coefficient is greater than zero, but it is closer to 0 than it is

to 1 => The positive relationship is very weak and likely unimportant.

    2. <u>Covariance</u> for <mark>Total Participation and GDP per capita</mark> is =

➤ cov(Total,GDPpercap) = 68890.42

Meaning: Covariance > 1 => These two variables are directly proportional.

    <u>Correlation Coefficient</u> for <mark>Total Participation and GDP per capita</mark> is =
➤ cor(GDPpercap, Total) = 0.7325252.

Meaning: Correlation Coefficient is greater than zero and it is closer to 1 than it is to 0 => The positive relationship is strong and remarkable.

For the satisfaction of this question, correlation coefficients must be considered and compared. Correlation coefficient value of Total Participation and GDP (0.13) is smaller (with huge difference) than of Total Participation and GDP per capita (0.73). This means that GDP per capita is more responsive to total participation rate than it is for GDP.

Why is that?

Higher ratio of working people will cause a higher GDP, but it doesn't always make the citizens richer. Nominal GDP itself may not always be the best measurement to measure a nation's prosperity. Calculating GDP per capita and comparing the results will give us better answers.

Population is a big factor for this. A country with large population may have a GDP of X, and another country with smaller population may have the same GDP of X as well, but per capita income will be different in these countries. GDP per capita for the first country will be less than the second country; meaning the people in first country are poorer.
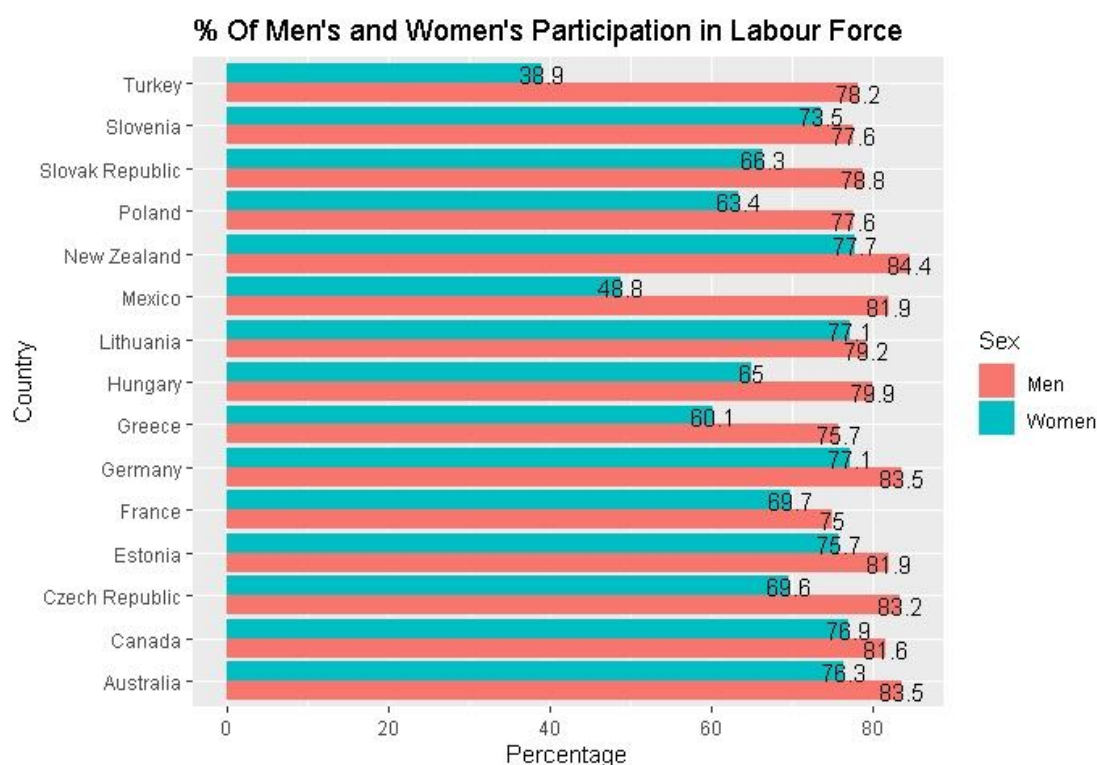
Therefore, again, GDP per capita will make us understand the concept of "prosperity of a nation's people" more correctly.

♦ Visual Summary of The Data

a.   Are men and women equally participated in labour force?

The bar chart below shows the percentage of participation in labour force of two genders in given countries.

➢ barwm <-ggplot(sdf,aes(Percentage,Country,fill=Sex)) + geom_bar(stat="identity",position = "dodge") + scale_fill_discrete(labels=c("Men","Women"))

➢ barwmyazili <- barwm + geom_text(aes(label = round(Percentage, 1)), position = position_dodge(0.9))

➢ barwmyazili+labs(title="% Of Men's and Women's Participation in Labour Force")

% Of Men's and Women's Participation in Labour Force

| Country	Women	Men
Turkey	38.9	78.2
Slovenia	73.5	77.6
Slovak Republic	66.3	78.8
Poland	63.4	77.6
New Zealand	77.7	84.4
Mexico	48.8	81.9
Lithuania	77.1	79.2
Hungary	65	79.9
Greece	60.1	75.7
Germany	77.1	83.5
France	69.7	75
Estonia	75.7	81.9
Czech Republic	69.6	83.2
Canada	76.9	81.6
Australia	76.3	83.5

As seen in the bar chart above, they are not. Men are more participated in labour force than women in every country. The difference between rates depend on the country.

There is no country that has equal ratios for men and women, but they are very close in Germany, Lithuania and Canada. Even though New Zealand has the highest rate for participation of women in labour force, with 77.7%; the difference with men is about 7%.

The smallest difference is in Lithuania with 2.1% while the greatest difference is in Turkey with 39.3%, which is even a larger number than the rate of women's participation itself, making it the country which has the smallest rate of women's participation.

b.   How does total participation in labour force effect the country's GDP? Does it affect GDP per capita more directly (relatively to other countries in the data list) than it does GDP?

➢ ggplot(participation_data,aes(Total,GDPpercap))+geom_smooth()+labs(x="Rate of total participation in labour force",y="GDP per capita",title="Relationship of total participation in labour force and GDP per capita(in US dollars)")
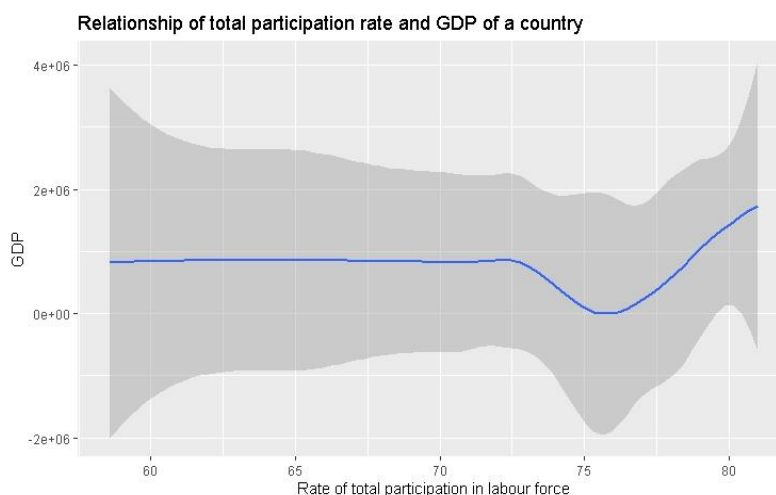
We see that there doesn't seem to be a positive or negative relationship between nominal GDP's of different countries compared with each other's participation rates. Reasons are as stated; a country's population is effective on this. Countries with smaller population couldn't have greater GDP than countries with big populations, even if their participation rate is high.

> ggplot(participation_data,aes(Total,GDP))+geom_smooth()+labs(x="Rate of total participation in labour force", title = "Relationship of total participation rate and GDP of a country")

Relationship of total participation in labour force and GDP per capita(in US dollars)
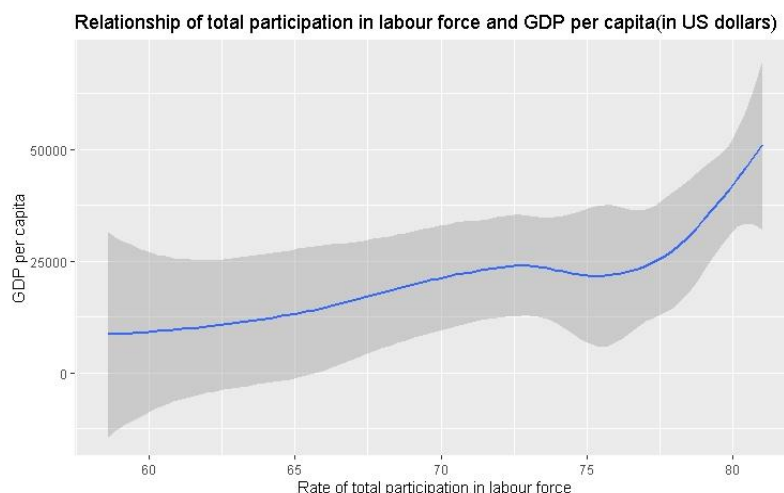

GDP per capita is a good measurement for this case. As seen in graph above, as participation rate increases; GDP per capita tends to increase as well.

♦ Analyzing the relationship between 2 variables

a.   Are men and women equally participated in labour force?

> ggplot(participation_data, aes(Men,Women,color=Country)) + geom_point() + lims(x=c(35,85),y=c(35,85)) + labs(x="% Of Men's Participation in Labour Force",y="% Of Women's Participation in Labour Force")
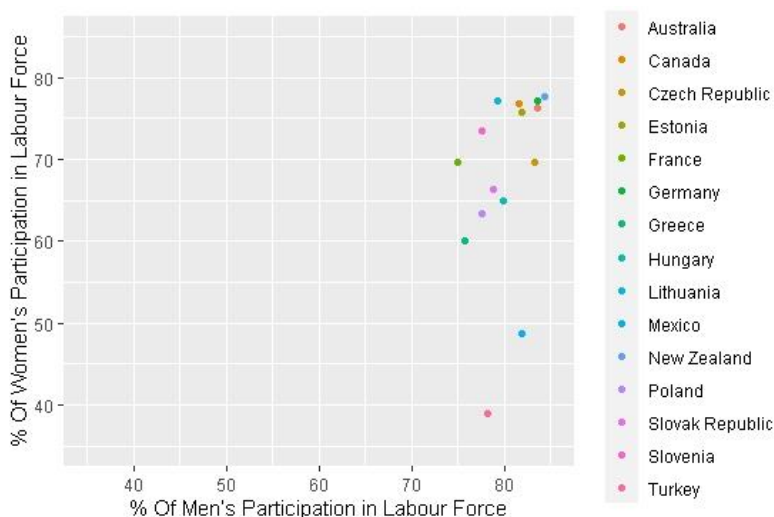
In the scatter plot above, the points are scattered around the ratios where men's are always high, while they are much more various for women. This shows us women's participation rates vary very differently upon the country.

This plot also tells us that there is a very weak correlation between these 2 variables. While men's rates dwell around 75% and 85%, women's are between <40% and <%80.

So, men are women are definitely not equally participated in the labour force.

➢ boxplot(Women,main="Boxplot for % Of Women's Participation in Labour Force")

**Boxplot for % Of Women's Participation in Labour Force**
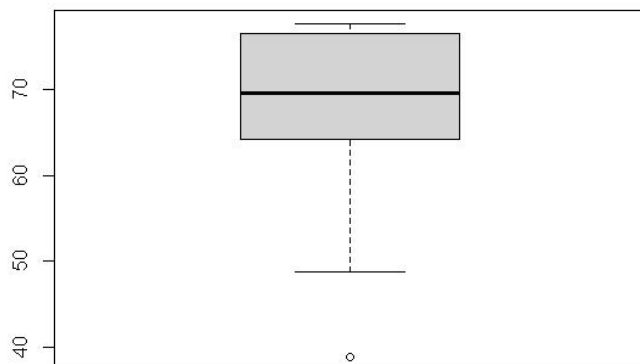
➢   The mean of women's participation rates is 67%. The minimum rate of participation is 38% while the maximum rate is 77%. The 1 outlier is the minimum value that is 38% which is Turkey's.

Absolute value of the difference between 3rd quartile and maximum value (1.1%) is smaller than the absolute difference between 1st quartile and the smallest value (25.3%). Which means that there are small numbers of countries where women's participation rate is high; and more countries have these rates for women relatively low.

➢ boxplot(Men,main="Boxplot for % Of Men's Participation in Labour Force")

**Boxplot for % Of Men's Participation in Labour Force**
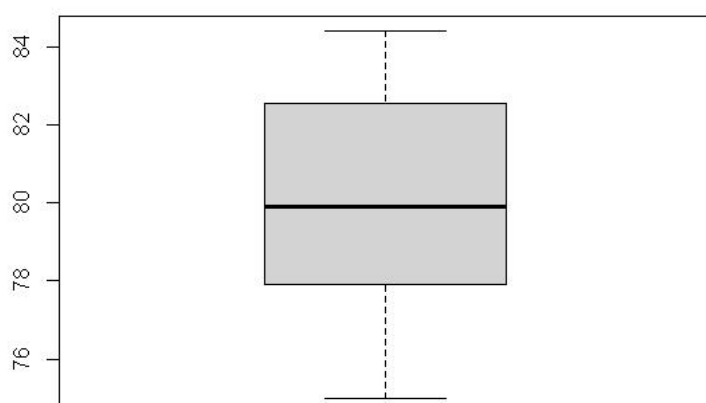
➢   The mean of men's participation rates is 80%. The minimum value is 75% (which is almost equal to the highest rate of women's!) while the maximum value is about 84%. There is no outlier.

The absolute value of the difference between 3rd quartile and maximum value is (2.42%) smaller than the absolute difference between 1st quartile and the smallest value (4.9%); but the difference is not so big. Meaning that men's participation rates are relatively close to each other.

b. **How does total participation in labour force effect the country's GDP? Does it affect GDP per capita more directly (relatively to other countries in the data list) than it does GDP?**

➢ ggplot(participation_data,aes(Total,GDP,color=Country))+geom_point()+labs(x="%O f Total Participation in Labour Force")

➢ ggplot(participation_data,aes(Total,GDPpercap,color=Country))+geom_point()+labs(x ="%Of Total Participation in Labour Force", y="GDP per capita (in US dollars)")
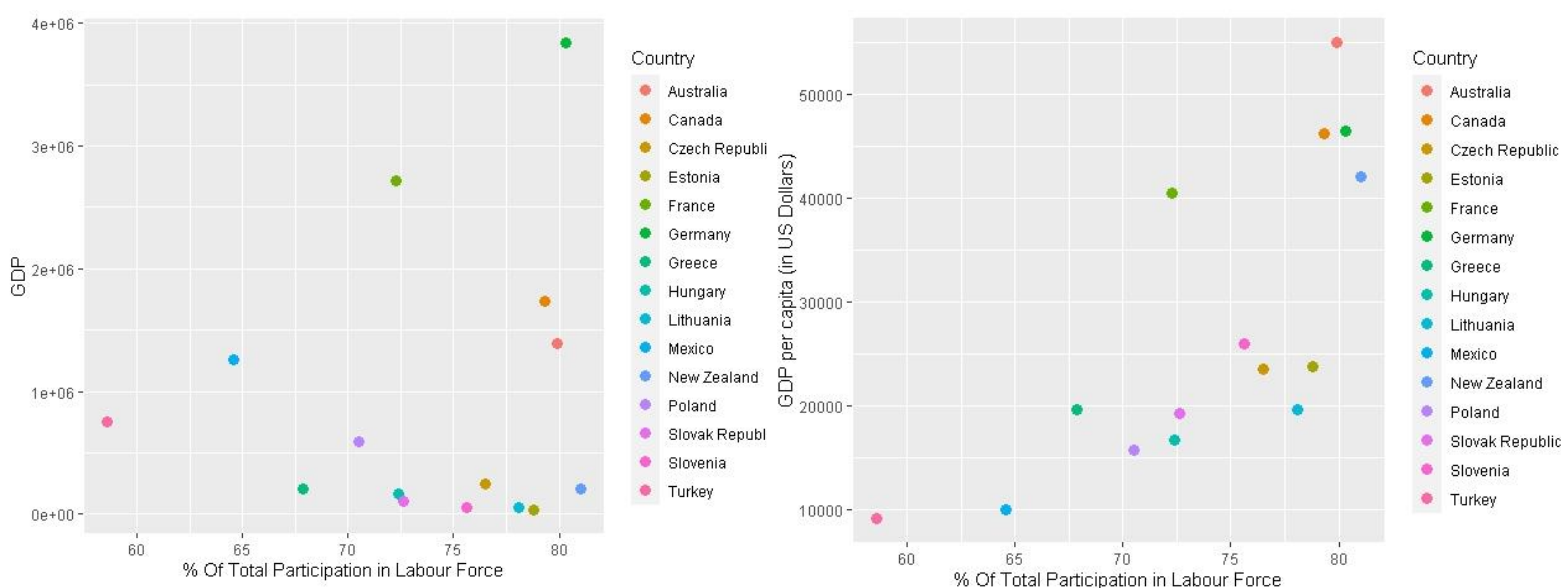
When these two scatterplots are compared, we see that the rate of total participation in labour force is more effective on GDP per capita rather than GDP itself. That is mostly related to the country's population. People of a country with large population would be considered less wealthy than a country that has a smaller population and the same GDP.

GDP per capita is a better way to measure the country's prosperity.

QUESTION 2

Research question: "***What kind of characteristics do students with the higher GPA have?***"

Observation 1: Correlation between GPA and sleeping hours of students is: **- 0.059**, which means that more successful students tend to sleep slightly less than less successful ones. The negative relationship is not very strong, but it is unignorable.
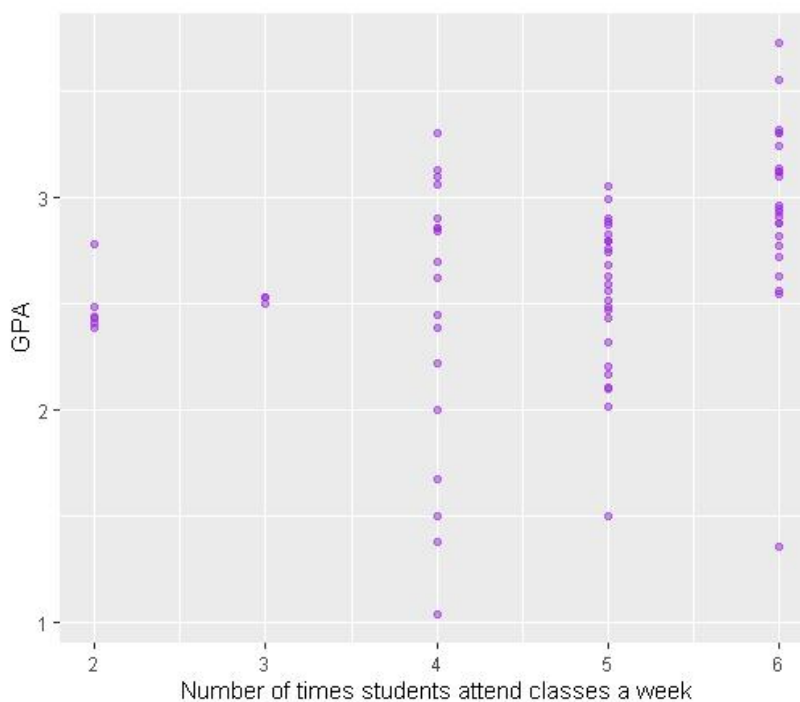
➢ ggplot(gpa2020,aes(sleep,gpa))+geom_point(color="red",alpha=.5)+labs(x="Number of hours that students sleep",y="GPA")
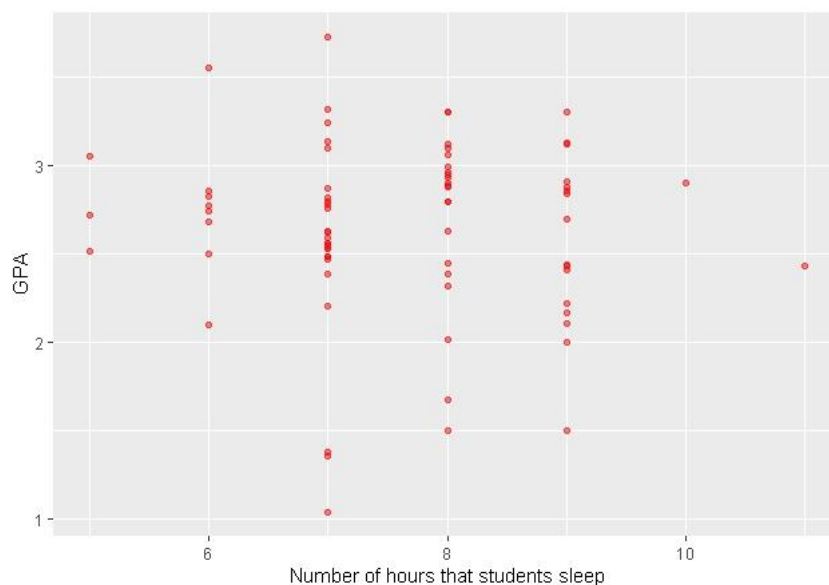
**Observation 2:** Students who attend the classes more, tend to have higher GPA. Correlation between these two variables is: **0.31**.

➢ ggplot(gpa2020, aes(gpa,attend)) + geom_point(color="purple",alpha=.5) + labs(x="GPA",y="Number of times students attend classes a week")

.**Observation 3:** Students with higher GPA study more days before the exam. The correlation between GPA and number of days students study for exam is: **0.24**.Most students study 7 days before the exam

➤ ggplot(gpa2020,aes(daysbefore,gpa))+geom_point(color="red",alpha=.5)+labs(x="Number of days students study for exam",y="GPA")
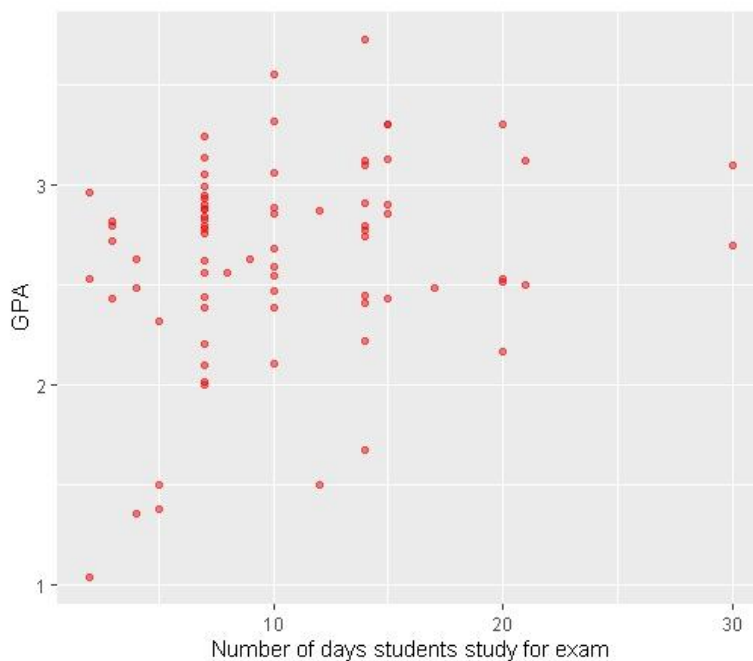

Observation 4: Taking notes during the lecture does not affect the grades of a student. That's mostly because during the pandemic, the lectures are done online, and lecturers upload the notes to the moodle. Students don't need to take notes, they study the notes after the lecture.

➤ ggplot(yesno,aes(yn,gpa))+ _point(alpha=.2,color="blue") + labs(x="Yes/No",y="GPA") +ggtitle("Do students take notes during the lecture, and how does in affect their grades?")
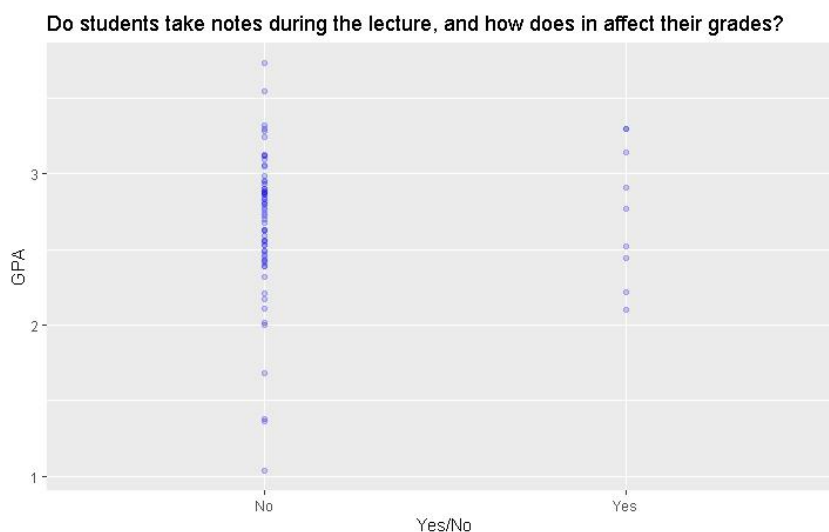
➤ ggplot(yesno,aes(yn,gpa))+geom_boxplot()+labs(x="Yes/No",y="GPA")+ggtitle("Do students take notes during the lecture, and how does in affect their grades?")

Do students take notes during the lecture, and how does in affect their grades?
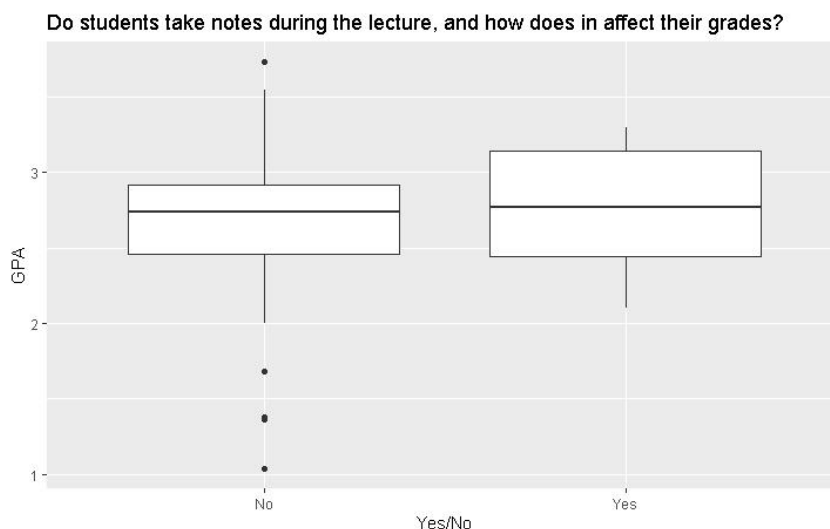
It can also be seen in this boxplot, most of the students don't take notes; and there is not really a considerable difference of GPA (mean, 1st and 3rd quartiles) between those who do and who don't.

Both of the students with highest and lowest GPAs don't take notes during the lecture.

Observation 5: Students whose GPA is equal to or greater than 3 study at least 7 days before an exam.

Those who study 30 days obviously study regularly every day, regardless of any exams. Many students study 7 days before an exam.

➤ ggplot(gpa2020,aes(daysbefore,gpa))+geom_point(color="red",alpha=.5)+labs(x="Number of days students study for exam",y="GPA")
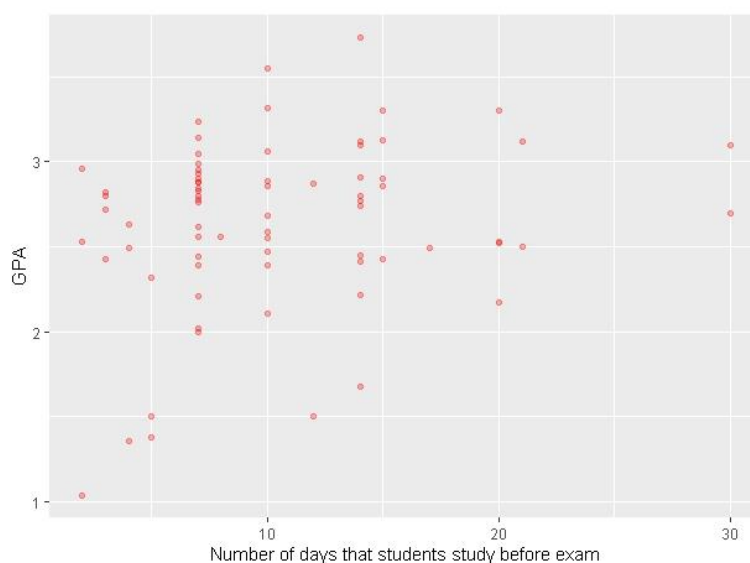
Observation 6: It appears that being prepared for lessons doesn't really make a difference; checking the boxplot here, we see that the student with highest GPA is neutral about it. We also see that means and quartiles of these boxplots are very close to each other, so GPA (or student's academical success) doesn't depend on preperation for lectures.

➢ ggplot(good,aes(importance,gpa))+geom_boxplot()+labs(x="Importance of being prapered for lessons, according to students",y="GPA")