

**ANL501**

**Data Visualisation and Storytelling**

**Tutor-Marked Assignment July 2024 Presentation**

Name: Que Ming Hui, Max

Submission Date: 08 September 2024

**Question 1 (a)**

**Introduction**

Work forms a key component of our lives and it is commonly associated with our level of happiness. To what extent, however, does work shape happiness? How much do working conditions impact happiness around the world? This report seeks to answer these questions and find out the relationship between work and happiness.

**Data Preparation**

The following datasets were used for analysis in this report: ANL501\_TMA\_JUL24\_HOURS.xlsx, ANL501\_TMA\_JUL24\_WORKING.xlsx and ANL501\_TMA\_JUL24\_HAPPINESS.csv. In addition, a dataset on the mean weekly hours worked per employed person by sex and occupation in each country, obtained from the International Labour Organization (ILO) (https://rplumber.ilo.org/data/indicator/?id=HOW\_TEMP\_SEX\_OCU\_NB\_A&type=label&format=.csv), was also used in this report.

The first dataset, ANL501\_TMA\_JUL24\_HOURS.xlsx, contains the average number of working hours per week in OECD (Organisation for Economic Co-operation and Development) countries and Singapore. This is an Excel workbook with 4 sheets: the first 3 sheets containing the working hours data from OECD countries (from the total “OECD Total”, male “OECD Male” and female “OECD Female” populations respectively) from 2010 to 2022, and the last sheet containing data from the total Singapore population from 2013 to 2023.

The data from “OECD Total” and “Singapore Total” sheets were imported first so that comparisons between the working hours from the total population of each country can be made:

#import working hours (total) for OECD countries and Singapore

oecdtotalhours <- read\_excel("ANL501\_TMA\_JUL24\_HOURS.xlsx", sheet = "OECD Total")

sghours <- read\_excel("ANL501\_TMA\_JUL24\_HOURS.xlsx", sheet = "Singapore Total")

Data profiling on “OECD Total” was performed using the str()and head()functions in R:

str(oecdtotalhours)

head(oecdtotalhours) %>%

print(width=Inf)

From this, it was observed the data from “OECD Total” is arranged in the wide format with each year as a variable on each column. To make this dataset more suitable for merging and comparing with other datasets, the “OECD Total” data frame was pivoted to a long format:

pivottotalhours <- oecdtotalhours %>%

pivot\_longer(!Country,names\_to="year",values\_to = "hours")

pivottotalhours

This reshaped the “OECD Total” data to a long format with 3 columns: Country, year and hours (where ‘hours’ represent the average weekly working hours for each country in each year). However, it was noted that the observations under the ‘year’ column of the reshaped data was converted to a character data type. This was resolved by changing it back to an integer type:

pivottotalhours$year <- as.integer(pivottotalhours$year)

pivottotalhours

The reshaped “OECD Total” data was also checked for any missing observations:

print(pivottotalhours[!complete.cases(pivottotalhours),], n=Inf)

Next, data from the “Singapore Total” sheet was similarly examined with the str()and head()functions. The data here is also arranged in a wide format in the same manner as “OECD Total”.

In order to use the “Singapore Total” data for analysis with other datasets, the first row containing annual average working hours per week shall be extracted. Before extracting, the column “Variable” was renamed to “Country”. The first row was then extracted and the first value of that row was renamed from “AnnualAverage” to “Singapore”. These steps were performed so that the column and row names align with other datasets.

#rename the first column to Country to align with columns in OECD Total

names(sghours)[names(sghours) == "Variable"] <- "Country"

#extracts the Annual Average row in Singapore Total and renames the first value to Singapore as the row name

sgannualhours <- sghours[1,]

sgannualhours[1] <- "Singapore"

It was also observed that the observations under the “2015” column of the “Singapore Total” data were character types. They were hence converted to a numeric type:

sgannualhours$"2015" <- as.numeric(sgannualhours$"2015")

sgannualhours

The extracted first row of the “Singapore Total” data was then similarly reshaped to the long format, followed by converting the observations in the ‘year’ column from character to integer type, and then checked for missing values:

#pivots sgannualhours from wide to long

pivotsghours <- sgannualhours %>%

pivot\_longer(!Country,names\_to="year",values\_to = "hours")

#converts all Year entries from character to numerical (integer)

pivotsghours$year <- as.integer(pivotsghours$year)

#filters out rows with missing values

pivotsghours[!complete.cases(pivotsghours),]

When converting the year values to integer type, the character ‘2002’ was converted to ‘NA’ instead. The value ‘2022’ was introduced back to fix this issue:

pivotsghours[is.na(pivotsghours$year),"year"] <- 2022

pivotsghours

The data from “OECD Total” and “Singapore Total” can now be joined in one data frame “workhours”.

#merges data from OECD Total and Singapore Total into one data frame

workhours <- full\_join(pivottotalhours,pivotsghours)

In the next step on the preparation of the working hours dataset, a new column in this “workhours” data frame was created containing the country codes of each country. The reason for creating this column is because the names of the countries in this dataset may not match with the country names in the other datasets (for example, the Republic of Korea may be named “Korea” in this dataset but may be called “South Korea” in another dataset). This presents an issue when attempting to merge two datasets as countries with different names in the datasets may be unknowingly left out and data from these countries will be missing in the merged dataset.

To solve this issue, the R package “countrycode” was utilized. This package is able to convert a country name to a specified standardized country code. This ensures all datasets share the same country codes and they can be merged without the problem of different naming conventions of each country.

Using this “countrycode” package, the following column was created for the “workhours” data frame which generated the country codes for each country:

#creates a new column for country codes. Here the country names are converted to GENC 3-letter codes

workhours <- workhours %>%

mutate(country.code = countrycode(workhours$Country,origin = 'country.name', destination = 'genc3c'))

Next, the “workhours” data frame was checked for any missing values:

#checks for missing values in "workhours" data frame

workhours[!complete.cases(workhours),]

It was noted that there was no data on the working hours in South Korea from ANL501\_TMA\_JUL24\_HOURS.xlsx. Thus, the dataset from ILO was used to fill the missing data on South Korea in the “workhours” data frame. This ILO dataset contains the mean weekly working hours of each country, categorised by sex and occupation. The working hours data in South Korea, based on the total population (including all sexes and occupation types) and from 2010 to 2022, was extracted from the ILO dataset and then input into the “workhours” data frame:

#filters out data on South Korea from the ILO dataset (stored in "countrytotalhours" variable) and saved as "koreahours" variable

koreahours <- filter(countrytotalhours, country.code == "KOR",

sex.label == "Sex: Total", classif1.label == "Occupation (ISCO-08): Total") %>%

subset(select=c("Country","year","hours")) %>%

arrange(year)

#missing data on working hours in South Korea is filled by the "koreahours" data, which itself is from the ILO dataset

workhours[workhours$country.code == "KOR","hours"] <- subset(filter(koreahours, year %in% c(2010:2022)), select = "hours")

There were also no data on the working hours in Canada and Japan in ANL501\_TMA\_JUL24\_HOURS.xlsx, despite being OECD countries. Similarly, data on the working hours in these two countries from the ILO dataset were extracted and inserted into the “workhours” data frame. An example is shown below for Canada:

#filters out data on Canada from the ILO dataset and saved as "canadahours" variable

canadahours <- filter(countrytotalhours, country.code == "CAN",

sex.label == "Sex: Total", classif1.label == "Occupation (ISCO-88): Total") %>%

subset(select=c("Country","year","hours","country.code")) %>%

arrange(year)

#adding rows on working hours in Canada to "workhours" data frame

workhours <- rbind(workhours,canadahours[canadahours$year %in% c(2010:2022),])

The next dataset, ANL501\_TMA\_JUL24\_WORKING.xlsx, contains information on working conditions (paid annual leave, severance pay and unemployment protection) from each country, from 2004 to 2020.

This dataset was saved as a variable “workcond” and the column names were changed to shorter names so that they were easier to read:

#import data on working conditions from ANL501\_TMA\_JUL24\_WORKING.xlsx

workcond <- read\_excel("ANL501\_TMA\_JUL24\_WORKING.xlsx")

#data profiling on workcond

str(workcond)

head(workcond)%>%

print(width=Inf)

#renaming column names to shorter ones

colnames(workcond) <- c("Country","Economy Code","year","leave\_1year","leave\_5\_ear","leave\_10year","leave\_avg","sevpay\_1year","sevpay\_5year","sevpay\_10year","sevpay\_avg","unemployment\_protection")

From the data profiling of “workcond” data frame, it was observed that the values under the severance pay columns, while resembling numerical values, were all of the character data types. These were subsequently changed to numerical type using the sapply() function. In addition, the values under the “year” column were changed from numerical to integer type.

#saves severance pay column names to a variable "workcondchar"

workcondchar <- c("sevpay\_1year", "sevpay\_5year", "sevpay\_10year", "sevpay\_avg")

#using sapply to convert values in severance pay columns to numeric type

workcond[workcondchar] <- sapply(workcond[workcondchar] ,as.numeric)

#converts values under the "year" column from numeric to integer type

workcond$year <- as.integer(workcond$year)

#checks structure of 'workcond' to ensure data types are correct

str(workcond)

A new column containing the country codes for each country was also created for the “workcond” data frame, before conducting sanity checks for any missing values:

#creates a new column for country codes

workcond <- workcond %>%

mutate(country.code = countrycode(workcond$Country,origin = 'country.name', destination = 'genc3c'))

#checks for any missing values in workcond

print(workcond[!complete.cases(workcond),], width=Inf, n=Inf)

The last dataset, ANL501\_TMA\_JUL24\_HAPPINESS.csv, contains the values of the Life Ladder, Log GDP (Gross Domestic Product) per capita, positive affect and negative affect of each country. This dataset is saved under the variable “happiness”.

Data profiling of “happiness” revealed the column name listing the countries as “Country name”. This was renamed to “Country” to align the column name to the previous datasets so that it could be merged later. The values under the “year” column were similarly converted to integer type. As was done for the previous datasets, a column on the country codes was also generated for the “happiness” data frame:

#saves ANL501\_TMA\_JUL24\_HAPPINESS.csv data to a variable

happiness <- read\_csv("ANL501\_TMA\_JUL24\_HAPPINESS.csv")

#data profiling of 'happiness' data frame

str(happiness)

head(happiness)%>%

print(width=Inf)

#renaming the column name from 'Country name' to 'Country'

names(happiness)[names(happiness) == "Country name"] <- "Country"

#converts values under the year column from numeric to integer

happiness$year <- as.integer(happiness$year)

#creates a new column for country codes

happiness <- happiness %>%

mutate(country.code = countrycode(happiness$Country,origin = 'country.name', destination = 'genc3c'))

#checks for missing values in 'happiness'

print(happiness[!complete.cases(happiness),], width=Inf, n=Inf)

#it is noted that there is no country code for Somaliland region in the "happiness" data frame.

#and that country code is assigned as "NA". This is not an issue as there is no data on Somaliland region in the other datasets

#so there is no concern when this dataset is joined to other datasets by the country codes.

**Analysis and Discussion**

**GDP Per Capita**

The Life Ladder is a measurement of life satisfaction and a general indicator of happiness. Looking at trends in Life Ladder values among countries could reveal some insight on the relationship between work and happiness.

Firstly, a function “lifeladderfun” was created in R to filter out countries with either the top ‘x’ or bottom ‘x’ values in Life Ladder, where ‘x’ is a numerical user input. This function calculates the mean Life Ladder from 2018 to 2023 for each country, and selects the countries with top ‘x’ or bottom ‘x’ mean Life Ladder, depending on the user input:

#this creates a function that first groups 'happiness' data by country, then aggregate it by mean life ladder,

#followed by taking the top or bottom x of the mean values (depending on the input) and

#finally this function filters out countries that match the country names that are top or bottom x in life ladder values

lifeladderfun <- function(x, ladderstats="max1"){

if(ladderstats== "max1"){

topladderdf <- happiness %>% group\_by(`Country`) %>%

summarise(avglifeladder = mean(`Life Ladder`, na.rm = TRUE)) %>%

slice\_max(avglifeladder, n = x)

filter(happiness, `Country` %in% topladderdf$`Country`)

} else if(ladderstats== "min1"){

bottomladderdf <- happiness %>% group\_by(`Country`) %>%

summarise(avglifeladder = mean(`Life Ladder`, na.rm = TRUE)) %>%

slice\_min(avglifeladder, n = x)

filter(happiness, `Country` %in% bottomladderdf$`Country`)

}

}

With this, a plot of Life Ladder over the years can be constructed, with countries having the top 10 and bottom 10 mean Life Ladder highlighted:

#p1 generates a line plot of the Life Ladder over the years for each country

p1 <- ggplot(mapping=aes(x=year,y=`Life Ladder`)) +

geom\_line(data=happiness, aes(group=`Country`), colour = alpha("grey", 0.7)) +

geom\_line(aes(colour = `Country`), data = lifeladderfun(10,"max1")) +

geom\_line(aes(colour = `Country`), data = lifeladderfun(10,"min1")) +

guides(color = "none")

#q1 creates labels for the top and bottom countries in mean Life Ladder

#"laddergroup" data frame consists only of both top 10 and bottom 10 countries in Life Ladder from the 'happiness' data frame

#and "ladder\_label" is a column in "laddergroup" that consists of these country labels

q1 <- geom\_text\_repel(aes(colour=Country,label=ladder\_label),

data = laddergroup,size=5, direction="y",xlim = c(2024, NA),

hjust=-0.01,segment.linetype = "dotted")

p1 + q1 + expand\_limits(x=c(2005,2027)) +

scale\_x\_continuous(breaks = c(2005,2008,2011,2014,2017,2020,2023),

labels= c("2005","2008","2011","2014","2017","2020","2023")) +

labs(x = "Year",

title = "Life Ladder of each country from 2005 to 2023"

) +

theme(plot.title = element\_text(size = 18),

axis.text=element\_text(size=16),

axis.title=element\_text(size=16))

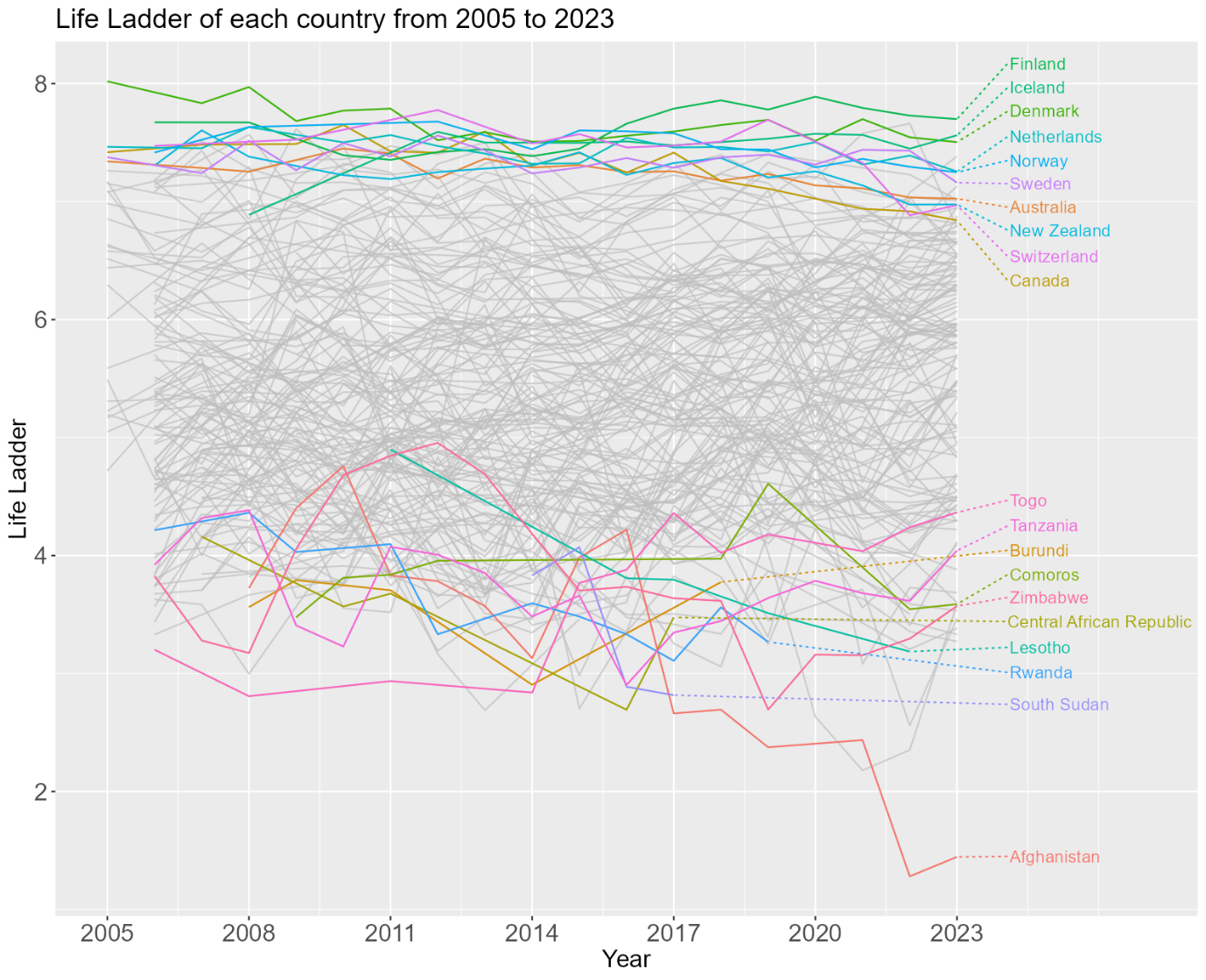


Figure 1. Plot of Life Ladder of each country over the years from 2005 to 2023. The top 10 countries and bottom 10 countries with the average Life Ladder over this year range are highlighted.

Based on Figure 1, the top 10 countries in Life Ladder are all members of OECD, with all 5 Scandinavian countries (Denmark, Finland, Iceland, Norway, Sweden) being in this top list. The bottom 10 countries in Life Ladder, on the other hand, are mostly from Africa.

The log GDP per capita, which can be used as an indirect indicator of per capita income (*Glossary | DataBank*, n.d.), from each of these top 10 and bottom 10 countries were examined in a plot below:

#Generates a line plot of the Log GDP per Capita over the years for top and bottom countries in Life Ladder

p3 <- ggplot(data = laddergroup, mapping=aes(x=year,y=`Log GDP Per Capita`,colour=lifeladderscore)) +

geom\_line(aes(group=Country))

#Creates labels for the top and bottom 10 Life Ladder countries in the plot

q3 <- geom\_text\_repel(aes(colour=lifeladderscore,label=ladder\_label),

data = laddergroup,size=5, direction="y",xlim = c(2024, NA),

hjust=0,segment.linetype = "dotted",show.legend = F)

p3 + q3 + expand\_limits(x=c(2005,2027)) +

scale\_x\_continuous(breaks = c(2005,2008,2011,2014,2017,2020,2023),

labels= c("2005","2008","2011","2014","2017","2020","2023")) +

labs(x = "Year",

title = "Log GDP Capita for countries with the top and bottom 10 mean Life Ladder values", color = "Life Ladder"

) +

theme(plot.title = element\_text(size = 18),

axis.text=element\_text(size=16),

axis.title=element\_text(size=16),

legend.title=element\_text(size=14),

legend.text=element\_text(size=12))

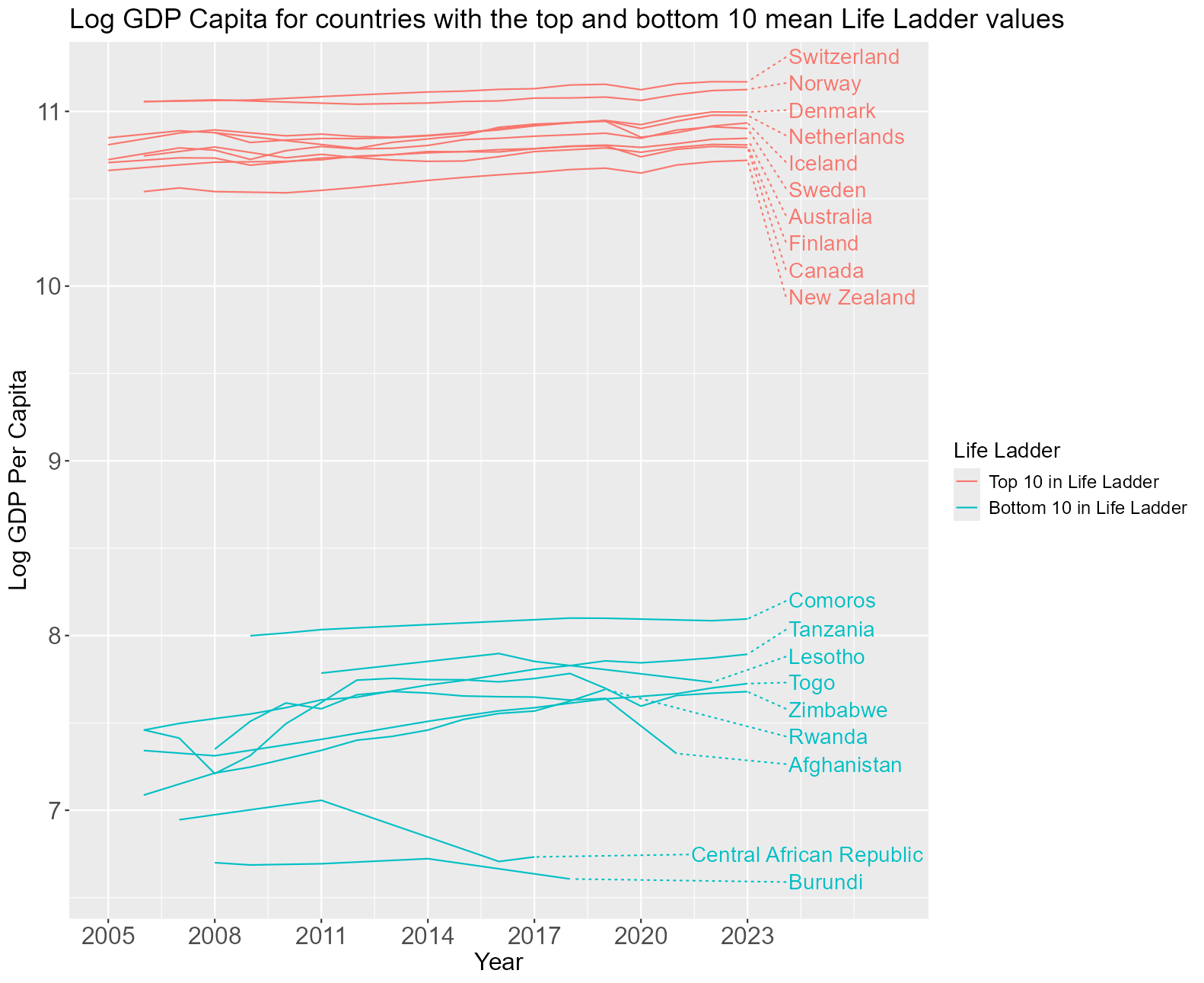


Figure 2. Plot of Log Gross Domestic Product (GDP) per capita, from 2005 to 2023, of the top 10 and bottom 10 countries in mean Life Ladder.

From Figure 2, we can see there is a wide gap in GDP per capita between the top 10 and bottom 10 countries in Life Ladder. This implies a relationship between economic prosperity and how satisfied one is in life. This relationship is more clearly seen from the scatterplot of Life Ladder against log GDP per capita for all countries:

#Generates a scatterplot of Life Ladder against Log GDP per capita

ggplot(data=happiness, aes(x=`Log GDP Per Capita`,y=`Life Ladder`)) +

geom\_point(alpha=0.5,show.legend = F) +

stat\_cor(p.accuracy = 0.001, r.accuracy = 0.01,size=6) +

geom\_smooth(colour="red",method="lm",fill=NA) +

labs(title = "Life Ladder against Log GDP Per Capita") +

theme(plot.title = element\_text(size = 18),

axis.text=element\_text(size=16),

axis.title=element\_text(size=16))

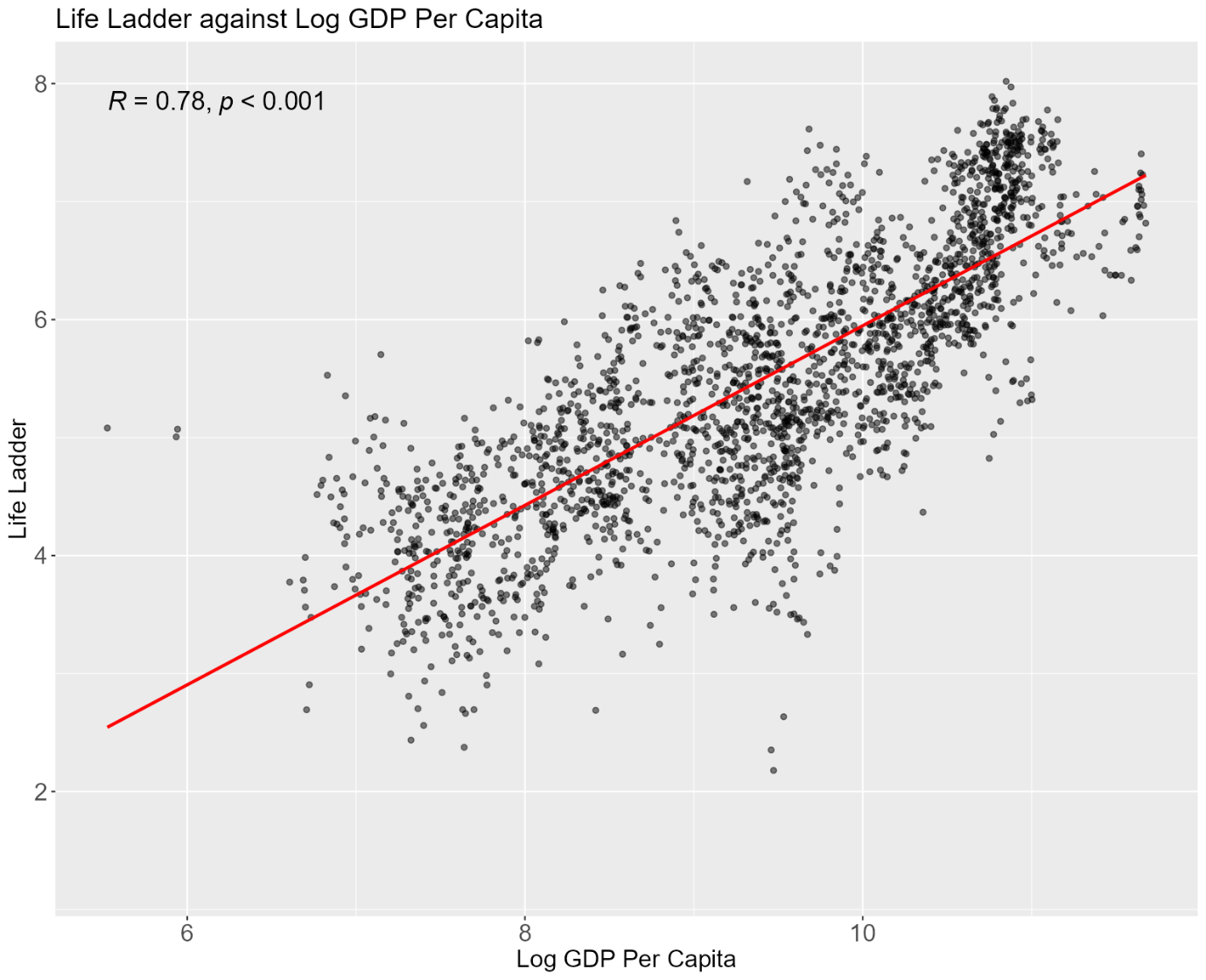


Figure 3. Plot of Life Ladder against log GDP per capita of each country between 2005 and 2023. A linear regression line is also plotted to show the correlation between Life Ladder and log GDP per capita.

A strong correlation between Life Ladder and log GDP per capita is observed in Figure 3. This provides further evidence that people tend to feel more satisfied in life when they have more income.

However, such relationship between the positive affect and log GDP per capita is not as significant.

#Generates a scatterplot of positive affect against Log GDP per capita

ggplot(data=happiness, aes(x=`Log GDP Per Capita`,y=`Positive Affect`)) +

geom\_point(alpha=0.5,show.legend = F) +

stat\_cor(p.accuracy = 0.001, r.accuracy = 0.01,size=6) +

geom\_smooth(colour="red",method="lm",fill=NA) +

labs(title = "Positive Affect against Log GDP Per Capita")+

theme(plot.title = element\_text(size = 18),

axis.text=element\_text(size=16),

axis.title=element\_text(size=16))



Figure 4. Plot of positive affect against log GDP per capita of each country between 2005 and 2023, with linear regression applied.

As seen in Figure 4, the correlation between positive affect and log GDP per capita is weak. Therefore, while a person’s life satisfaction is greatly affected by income, it does not have much impact on how much positive emotions a person experiences. This suggests that how much one earns from work may influence happiness in terms of life satisfaction but has little outcome on experiencing positive emotions.

**Annual Leave and Unemployment Benefits**

Next, work compensation of each country — in terms of paid annual leave and unemployment protection — are examined using the “workcond” dataset. The following is a graph of the average annual leave of the countries in the top 10 and bottom 10 Life Ladder over the years.

#p4 generates a line plot of paid annual leave over the years for top and bottom countries in Life Ladder

#"workcondgroup" data frame consists only of both top 10 and bottom 10 countries in Life Ladder from the 'workcond' data frame

p4 <- ggplot(data = workcondgroup, mapping=aes(x=year,y=`leave\_avg`,colour=lifeladderscore)) +

geom\_line(aes(group = `Country`))

#q4 creates labels for the top and bottom countries in mean Life Ladder

#"leave\_label" is a column in "workcondgroup" that consists of these country labels

q4 <- geom\_text\_repel(aes(colour=lifeladderscore,label=leave\_label),

data = workcondgroup,size=5, direction="y",xlim = c(2021, NA),

hjust=0,segment.linetype = "dotted",max.overlaps = 20, show.legend = F)

p4 + q4 + expand\_limits(x=c(2004,2026)) +

scale\_x\_continuous(breaks = c(2004,2008,2012,2016,2020),

labels= c("2004","2008","2012","2016","2020")) +

labs(x = "Year", y = "Number of days of paid annual leave",

title = "Average paid annual leave for countries with the top and bottom 10 mean Life Ladder values", color="Life Ladder"

) +

theme(plot.title = element\_text(size = 18),

axis.text=element\_text(size=16),

axis.title=element\_text(size=16),

legend.title=element\_text(size=14),

legend.text=element\_text(size=12))

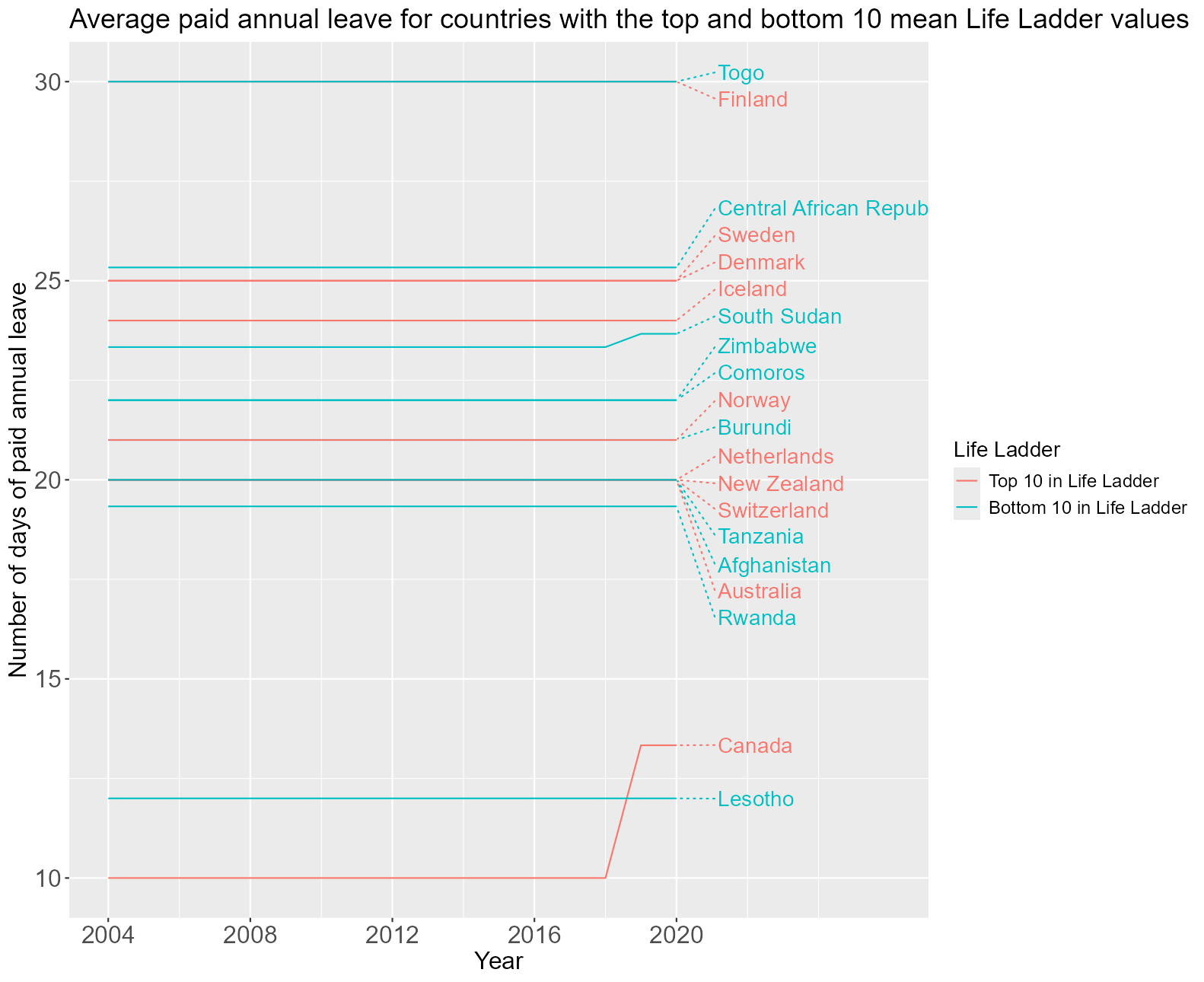


Figure 5. A graph showing the number of paid annual leave days from 2004 to 2020, for countries with the top 10 and bottom 10 Life Ladder.

In Figure 5, there is no discernible pattern between the countries with the top or bottom Life Ladder and the number of days of paid annual leave. There were countries with more days of paid annual leave but low Life Ladder (e.g. Togo and Central African Republic) and countries with fewer paid annual leave but high life Ladder (e.g. Canada).

Comparing this plot with a plot of the Life Ladder over the years for these countries in Figure 6, we can see while the number of days of annual leave remained almost unchanged from 2008 to 2020, there were fluctuations in the Life Ladder over the years, suggesting the number of paid annual leave days does not have an effect on the Life Ladder.

#p5 plots Life Ladder over the years

#"laddergroupfiltered" consists of data on life ladder from 2008 to 2020 for the countries with top and bottom 10 Life Ladder from the "happiness" data frame

p5 <- ggplot(mapping=aes(x=year,y=`Life Ladder`)) +

geom\_line(aes(colour = `Country`), data = laddergroupfiltered) #+

plot1 <- p5 + xlim(2007,2020) +

labs(x = "Year",

title = "Top 10 and bottom 10 countries in Life Ladder",tag = "A"

) +

theme(plot.title = element\_text(size = 16),

axis.text=element\_text(size=14),

axis.title=element\_text(size=14),

plot.tag = element\_text(size=18))

#p6 plots annual leave days over the years

p6 <- ggplot(data = filter(workcondgroup,year %in% c(2008:2020)), mapping=aes(x=year,y=`leave\_avg`)) +

geom\_line(aes(colour = `Country`)) #+

plot2 <- p6 + xlim(2007,2020) +

labs(x = "Year", y = "Number of days of paid annual leave",

title = "Average paid annual leave for countries with the top and bottom 10 mean Life Ladder values",

tag = "B"

) +

theme(plot.title = element\_text(size = 16),

axis.text=element\_text(size=14),

axis.title=element\_text(size=14),

plot.tag = element\_text(size=18))

#combines the Life Ladder graph and Annual Leave plot in the same page

ggarrange(plot1,plot2,ncol = 1, nrow = 2,common.legend = T,legend = "bottom")

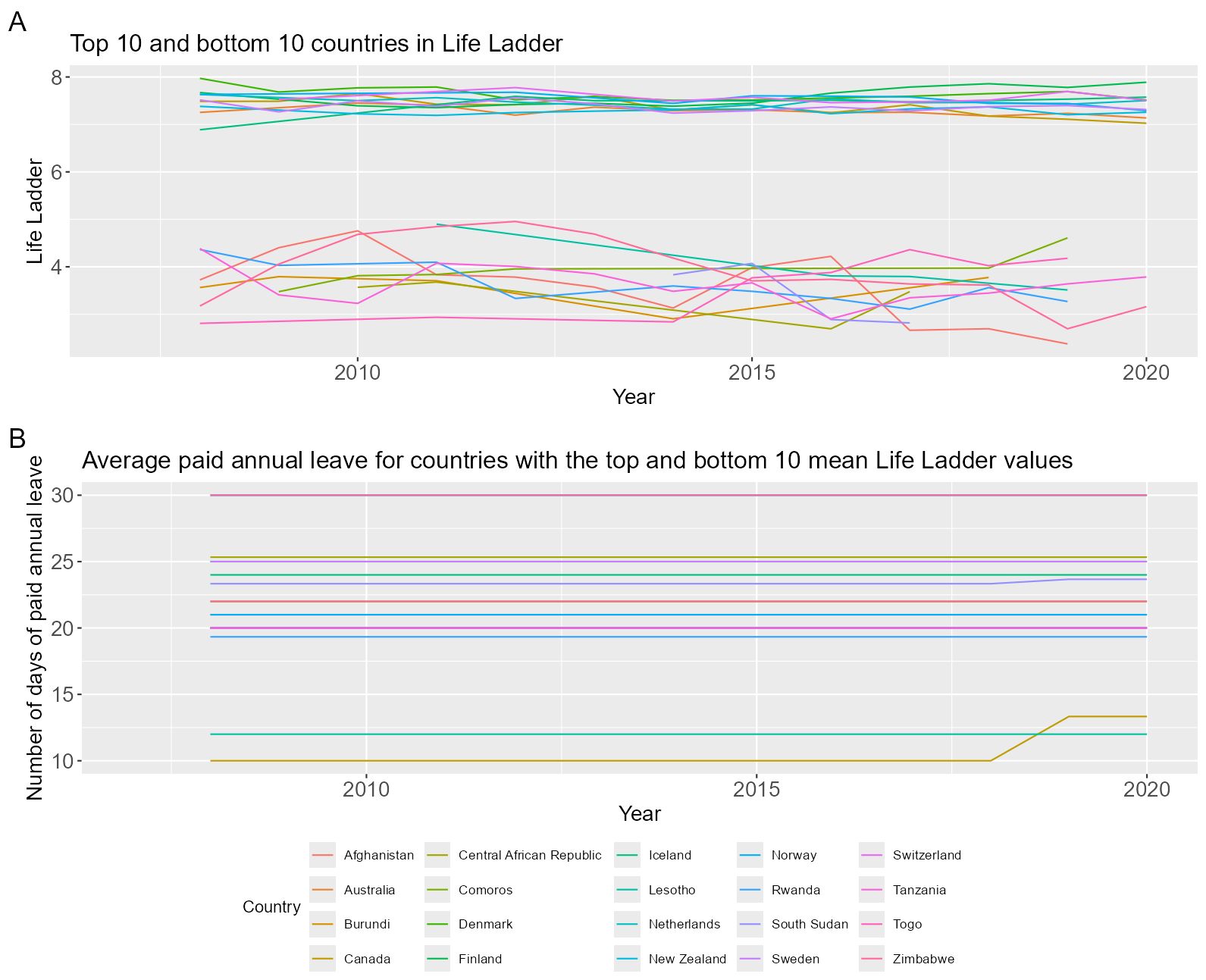


Figure 6. A) Life Ladder of countries that are top 10 and bottom 10 in mean Life Ladder. B) Average paid annual leave days of the countries with top 10 and bottom 10 mean Life Ladder. Both graphs are plotted against the year range from 2008 to 2020.

This is further supported by the scatterplot of Life Ladder against the number of days of paid annual leave of each country in Figure 7, in which no relationship between these 2 variables was identified. Hence, the number of paid annual leave days has little or no influence on happiness.

#merges working conditions dataset "workcond" with the "happiness" dataset into one data frame

#subset is used on 'happiness' data frame to remove the 'Country' column before joining

join2 <- left\_join(workcond,subset(happiness, select = -1),join\_by(country.code, year))

#Creates a scatterplot of Life Ladder against average annual leave

ggplot(data=join2, aes(x=leave\_avg,y=`Life Ladder`)) +

geom\_point(alpha=0.5) +

stat\_cor(p.accuracy = 0.001, r.accuracy = 0.01, size=6,label.x = 60,label.y = 7.5) +

geom\_smooth(colour="red",method="lm",fill=NA) +

labs(x = "Number of days of paid annual leave", y = "Life Ladder",

title = "Life Ladder against number of days of paid annual leave"

) +

theme(plot.title = element\_text(size = 18),

axis.text=element\_text(size=16),

axis.title=element\_text(size=16))

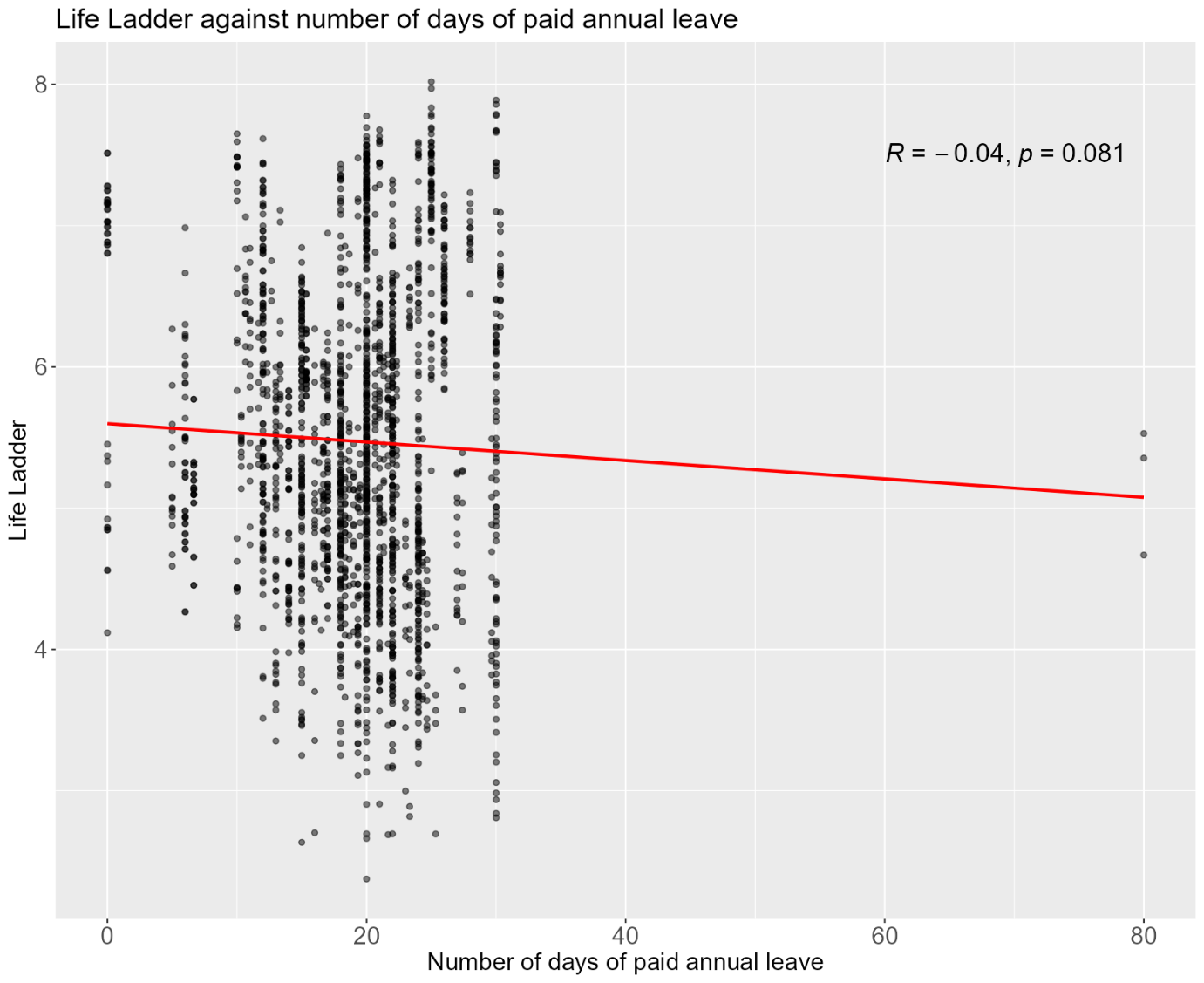


Figure 7. Plot of Life Ladder against average number of days of paid annual leave of each country. Linear regression is applied and it shows no significant relationship between the 2 variables.

Unemployment protection among the top 10 and bottom 10 countries in Life Ladder was also examined. Unemployment protection is defined as the eligibility of receiving unemployment benefits after one year of employment. The following table displays which of these countries have unemployment protection:

#creates table using gt package

#unemploy\_proc is a data frame that only consists of data on the unemployment protection status of each of the top 10 and bottom 10 Life Ladder countries

unemply\_proc %>%

gt() %>%

tab\_style(style = cell\_borders(sides = c("top","bottom"),

style ="hidden"),

locations = cells\_body(columns=unemployment\_protection,

rows = c(2:9,12:19))) %>%

tab\_style(style = cell\_borders(sides = "bottom",

style ="solid",

weight = px(1.5)),

locations = cells\_body(columns=c(Country,unemployment\_protection),

rows = 10)) %>%

tab\_style(style = list(cell\_fill(color="palegreen")),

locations = cells\_body(columns=unemployment\_protection,

rows = 1:10)) %>%

tab\_style(style = list(cell\_fill(color="lightcoral")),

locations = cells\_body(columns=unemployment\_protection,

rows = 11:20)) %>%

cols\_label(unemployment\_protection = "Unemployment Protection") %>%

cols\_align(align="center",columns = unemployment\_protection)

Table 1. List of countries with top 10 and bottom 10 Life Ladder, and whether they have unemployment protection based on the latest data (2020).



All top 10 countries in Life Ladder have unemployment protection and the opposite is true for countries at the bottom 10 in Life Ladder. While no concrete relationship between unemployment protection and happiness can be established from this observation, it is no doubt that this factor is present in countries which tend to be happy and may play a role in happiness.

**Working hours**

As previously noted, all countries in the top 10 Life Ladder are OECD members. The relationship between working hours and happiness in these countries, as well as other OECD countries, shall be explored. In addition, data on the working hours in Singapore was included in the analysis together with the rest of the OECD countries, due to similarities in economic developments between Singapore and OECD countries in terms of GDP per capita (OECD, 2022).

#declares the data, and maps the x and y variables to the graph

r1 <- ggplot(data=join1,aes(x=hours,y=`Life Ladder`))

#creates the points for the scatterplot

r2 <- r1 + geom\_point(data=filter(join1,!Country %in% c("Denmark","Finland","Iceland","Norway","Sweden","Singapore","Türkiye","Japan","Korea")),alpha=0.4,size=4) +

geom\_point(data=filter(join1,Country=="Singapore"),aes(colour=Country),alpha=0.7,size=4) +

geom\_point(data=filter(join1,Country %in% c("Denmark","Finland","Iceland","Norway","Sweden")),aes(colour=Country),alpha=0.7,size=4) +

geom\_point(data=filter(join1,Country=="Türkiye"),aes(colour=Country),alpha=0.7,size=4) +

geom\_point(data=filter(join1,Country=="Netherlands"),aes(colour=Country),alpha=0.7,size=4) +

geom\_point(data=filter(join1,Country=="Japan"),aes(colour=Country),alpha=0.7,size=4) +

geom\_point(data=filter(join1,Country=="Korea"),aes(colour=Country),alpha=0.7,size=4)

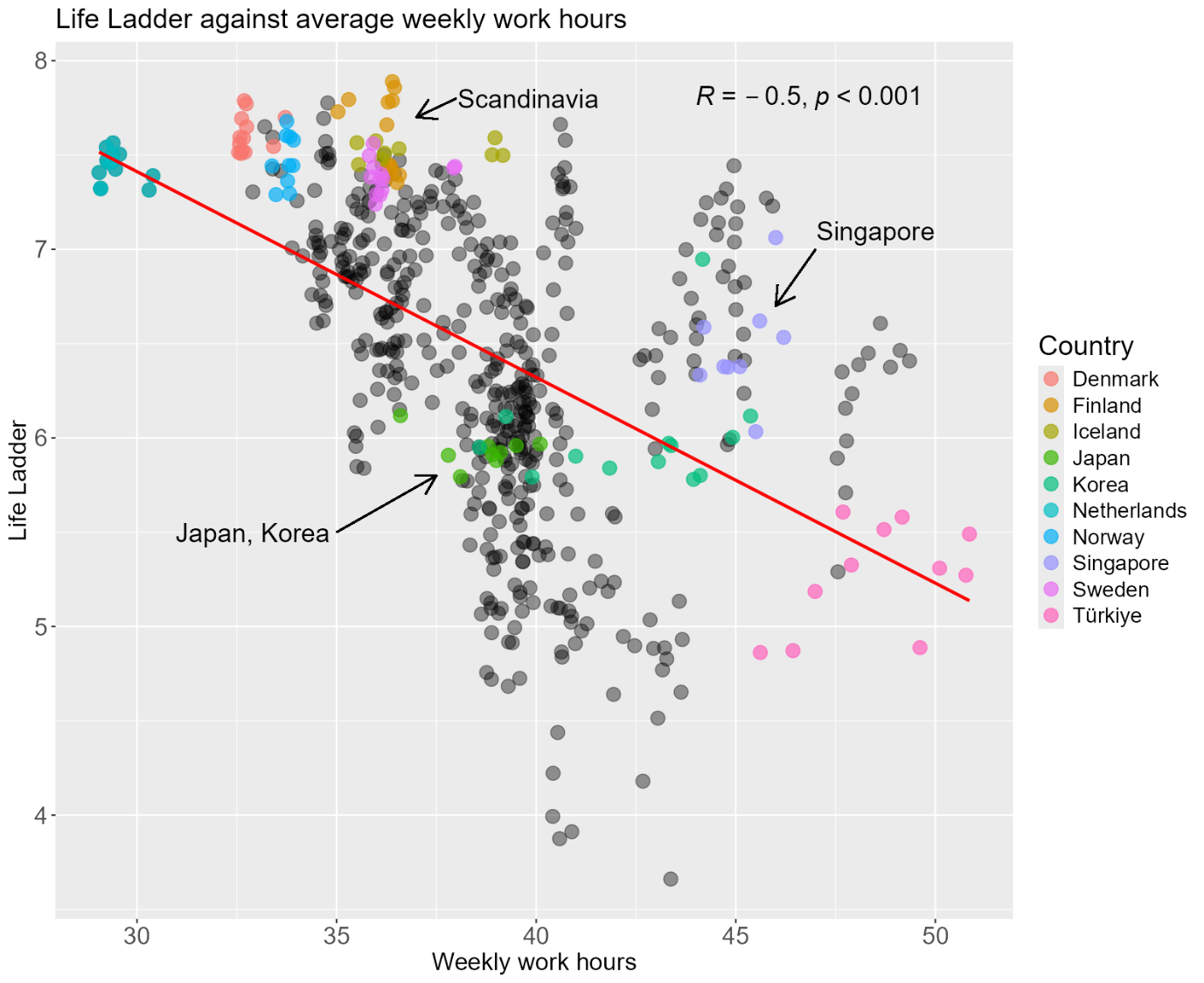


Figure 8. Plot of Life Ladder against average weekly work hours of OECD countries and Singapore between 2010 and 2022. The countries with highest and lowest working hours, Scandinavian countries and Singapore are highlighted in the figure.

The Life Ladder of the OECD countries and Singapore is plotted against the average weekly work hours in Figure 8. Netherlands, with the lowest work hours, had the one of the highest Life Ladder. In contrast, Türkiye had the highest work hours and while it did not have the lowest Life Ladder, its Life Ladder was still comparatively lower.

Singapore had one of the highest working hours and it was more than Japan and Korea. However, the Life Ladder in Singapore was still higher than in these two countries. This indicates that long working hours may not be the only contributors to perceived lower levels of happiness in Japan and Korea.

When compared with the Scandinavian countries which generally have lower working hours, the Scandinavians fared better in Life Ladder and had one of the highest Life Ladder.

Indeed, the figure shows moderate correlation between Life Ladder and work hours where life satisfaction generally increased as work hours got shorter. This suggests that one’s happiness can be significantly impacted by how long one has to work.

However, work hours alone does not influence happiness. While there is a correlation between these two factors, there are still observations of some OECD countries with shorter working hours but lower Life Ladder and vice versa. This implies there could be other elements that shape happiness.

**Conclusion**

Several work-related factors were explored on its effect on happiness. It was observed that countries with higher GDP per capita had higher Life Ladder. Life satisfaction is therefore greatly shaped by income, although positive affect was not influenced much by GDP per capita. This suggests that there are multiple aspects of happiness and how much one earns from work only affects a component that makes up happiness.

The number of days of paid annual leave had little or no outcome on happiness. However, it was observed the top 10 countries in Life Ladder had unemployment benefits, compared to the bottom 10 which had none.

There was also a correlation between the number of work hours and Life Ladder. In general, a person’s happiness level increases with shorter work hours. Scandinavian countries, popularly known for being the happiest countries, were showed to have the highest Life Ladder while having shorter working hours and also unemployment protection. This provided some justification that working conditions do have a significant influence on happiness.

However, there were some observations that do not fall into this general trend. While there was a relationship that shorter working hours led to higher life satisfaction, it does not necessarily hold true for all countries. For example, Singapore had longer working hours than Japan and Korea, which are sometimes known for their overwork culture, but yet Singapore scored a higher Life Ladder. There could be other elements that along with work, influence happiness. In addition, the dataset on working hours covered only OECD countries. The study on the connection between work hours and happiness could be further improved with more data from the rest of the countries. Based on the analysis conducted in this report, however, it is certain that work, being a key element in life, has a profound impact on happiness.

**References**

*Glossary | DataBank*. (n.d.). Retrieved August 19, 2024, from https://databank.worldbank.org/metadataglossary/statistical-capacity-indicators/series/5.51.01.10.gdp

OECD. (2022). *Society at a glance: Asia/pacific 2022* (p. 82). https://doi.org/10.1787/7ef894e5-en

*Statistics on working time*. (n.d.). ILOSTAT. Retrieved August 29, 2024, from https://ilostat.ilo.org/topics/working-time/

**Question 1 (b)**

Taking the plot of Life Ladder against the log GDP per capita in Figure 3, we can visualise the changes in the Life Ladder and GDP per capita of each country over the years in the form of an animated plot.

#gganimate package is used to create animations on ggplot

library(gganimate)

#viridis pacakage is loaded to use the viridis color map

library(viridis)

#creates an animated plot of Life Ladder against log GDP per capita

p7 <- ggplot(data=happiness, aes(x=`Log GDP Per Capita`,y=`Life Ladder`,colour = Country)) +

geom\_point(alpha=0.7,size=8,show.legend = F) +

theme(plot.title = element\_text(size = 18),

axis.text=element\_text(size=16),

axis.title=element\_text(size=16)) +

scale\_color\_viridis\_d() +

#shows the transitions in plot across the years

transition\_time(year) +

labs(title = "Year: {frame\_time}") +

shadow\_wake(wake\_length = 0.1, alpha = FALSE)

anim\_save("Figure 9.gif",p7)

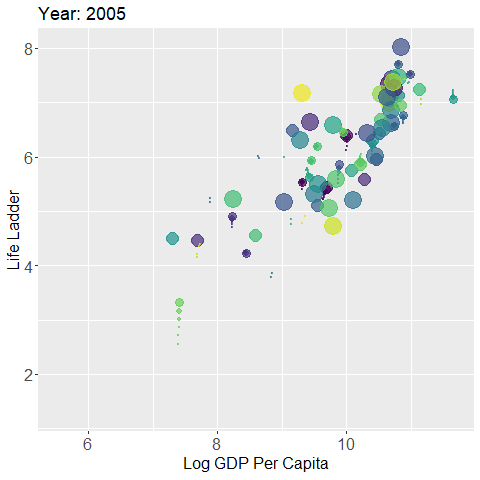


Figure 9. An animated plot of Life Ladder against log GDP per capita. This shows the transitions in the graph over the years from 2005 to 2023.

Based on the pattern observed in Figure 9, the log GDP per capita of each country generally followed a slight increasing trend while the Life Ladder fluctuated up and down over the years from 2005 to 2023. This may indicate the presence of other factors that influenced happiness in terms of life satisfaction.

However, countries that had higher GDP per capita over the years tend to continue have higher Life Ladder on average despite the fluctuations over time, as compared to countries with lower GDP per capita. Thus, this still corroborated with the observation from Figure 3 that the Life Ladder of the population in a country was significantly dictated by its GDP per capita.