

DS395B Final Project

Claudia Cabrera, Jimmy Horvath, Leyla Hatfield

2024-04-08

Load Libraries

```
library(tidyverse)
## └─ Attaching packages ─────────────────────────────────── tidyverse
  1.3.1 ─
  ##   ✓  ggplot2  3.4.4      ✓  purrr    1.0.2
  ##   ✓  tibble    3.2.1      ✓  dplyr    1.1.4
  ##   ✓  tidyr    1.3.0      ✓  stringr  1.5.0
  ##   ✓  readr    2.1.2      ✓  forcats  0.5.1
  ## └─ Conflicts ───────────────────────────────────
tidyverse_conflicts() ─
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()    masks stats::lag()
library(janitor)
##
## Attaching package: 'janitor'
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
library(countrycode)
library(stringr)
library(kableExtra)
##
## Attaching package: 'kableExtra'
## The following object is masked from 'package:dplyr':
##
##     group_rows
library(ggpirate)
library(ggpubr)
library(car)
## Loading required package: carData
##
## Attaching package: 'car'
## The following object is masked from 'package:dplyr':
##
##     recode
## The following object is masked from 'package:purrr':
##
##     some
library(scales)
##
## Attaching package: 'scales'
## The following object is masked from 'package:purrr':
##
##     discard
```

```

## The following object is masked from 'package:readr':
##
##     col_factor
library(stats)
library(modelsummary)

Import Data

vax <- read_csv("Covid Vax by region Clean (2).csv")
## Rows: 161 Columns: 4
## — Column specification
## Delimiter: ","
## chr (4): Country, Region, Partial, Full
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this
## message.
gdp <- read_csv("WorldbankGDP_data.csv")
## New names:
## Rows: 218 Columns: 7
## — Column specification
## _____ Delimiter: ","
chr
## (5): country, 2019, 2020, 2021, 2022 lgl (2): ...5, ...7
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this
## message.
## • `` -> `...5`
## • `` -> `...7`
life <- read_csv("WorldBank_LifeexpData.csv")
## New names:
## Rows: 218 Columns: 5
## — Column specification
## _____ Delimiter: ","
chr
## (4): country, 2019, 2020, 2021 lgl (1): ...5
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this
## message.
## • `` -> `...5`

```

Analyzing the impact of COVID-19 on Life expectancy

Research Questions:

1. Is there a significant difference between life expectancy before and after the pandemic? How did the impact of the pandemic on life exp vary by groups (region/ vax group/ gap group)?

- Are vaccination coverage and GDP good predictors of life expectancy post pandemic (2021).

Dataset Information [1](#)

GDP & Life expectancy data for 2019 and 2021 was sourced from the World Bank public use World Development Indicators dataset.

Clean Data

Vaccination Data

```
vax <- vax %>%
  clean_names() %>%
  mutate(partial = as.numeric(partial), full = as.numeric(full)) %>%
  mutate(region = as.factor(region), country = as.factor(country)))
## Warning: There were 2 warnings in `mutate()` .
## The first warning was:
## i In argument: `partial = as.numeric(partial)` .
## Caused by warning:
## ! NAs introduced by coercion
## i Run `dplyr::last_dplyr_warnings()` to see the 1 remaining warning.
# missing data
sum(is.na(vax))
## [1] 6
vax %>%
  filter(is.na(partial)) %>%
  kable() %>%
  kable_styling()


| country        | region        | partial | full |
|----------------|---------------|---------|------|
| French Guiana  | South America | NA      | NA   |
| Eritrea        | Africa        | NA      | NA   |
| Western Sahara | Africa        | NA      | NA   |


## remove missing data
vax_clean <- vax %>%
  filter(!is.na(partial))

# add in country code column for joining data later
vax_clean$country_code <- countrycode(vax_clean$country, origin =
"country.name", destination = 'iso3c')
## Warning in countrycode_convert(sourcevar = sourcevar, origin = origin,
destination = dest, : Some values were not matched unambiguously: Columbia,
Guniea, Kosovo
vax_clean <- vax_clean %>%
  mutate(country_code = ifelse(country == "Columbia", "COL",
                                ifelse(country == "Kosovo", "XKX",
                                      ifelse(country == "Guniea", "GIN",
country_code))))
## add in vax groups
```

```

vax_clean$vax_group <- cut(vax_clean$full,
                           breaks = quantile(as.numeric(vax_clean$full),
probs = c(0, 1/3, 2/3, 1), na.rm = TRUE),
                           include.lowest = TRUE,
                           labels = c("Low", "Medium", "High"))

rm(vax)

# View Data
head(vax_clean) %>%
  kable() %>%
  kable_styling() %>%
  scroll_box(width = "100%", height = "400px")

country    region    partial full country_code vax_group
Canada      North America   92 84 CAN      High
US          North America   80 68 USA      Medium
Mexico      North America   78 65 MEX      Medium
Guatemala   North America   54 43 GTM      Low
Belize       North America   65 57 BLZ      Medium
Honduras    North America   67 60 HND      Medium

```

Life Exp Data

```

life <- life %>%
  select(country, "2019", "2021") %>%
  rename(pre_lifeexp = "2019", post_lifeexp = "2021") %>%
  mutate(country = as.factor(country), pre_lifeexp = as.numeric(pre_lifeexp),
post_lifeexp = as.numeric(post_lifeexp))
## Warning: There were 2 warnings in `mutate()` .
## The first warning was:
## i In argument: `pre_lifeexp = as.numeric(pre_lifeexp)` .
## Caused by warning:
## ! NAs introduced by coercion
## i Run `dplyr:::last_dplyr_warnings()` to see the 1 remaining warning.
## add in country code variable
life$country_code <- countrycode(life$country, origin = "country.name",
destination = 'iso3c')
## Warning in countrycode_convert(sourcevar = sourcevar, origin = origin,
destination = dest, : Some values were not matched unambiguously: Channel
Islands, Created from: World Development Indicators
## Series : Life expectancy at birth, total (years)
##
##
##
## , Kosovo, Turkiye
# add in region
life <- life %>%
  mutate(country_code = ifelse(country == "Turkiye", "TUR",
                               ifelse(country == "Kosovo", "XKX",
country_code))) %>%
  mutate(country_code = case_when(country_code %in% vax_clean$country_code ~
country_code)) %>%

```

```

    filter(!is.na(country_code)) %>%
inner_join(vax_clean, join_by(country_code))

life_clean <- life %>%
  rename(country = country.x) %>%
  select(country, country_code, region, pre_lifeexp, post_lifeexp)

rm(life)

# View Data
head(life_clean) %>%
  kable() %>%
  kable_styling() %>%
  scroll_box(width = "100%", height = "400px")

country country_code region pre_lifeexp post_lifeexp
Afghanistan AFG Asia 63.6 62.0
Albania ALB Europe 79.3 76.5
Algeria DZA Africa 76.5 76.4
Angola AGO Africa 62.4 61.6
Argentina ARG South America 77.3 75.4
Armenia ARM Asia 75.4 72.0

```

GDP Data

```

gdp <- gdp %>%
  select(country, "2019", "2021") %>%
  rename(pre_gdp = "2019", post_gdp = "2021") %>%
  mutate(country = as.factor(country))

## format gdp value so can be changed to numeric var

gdp$pre_gdp <- str_replace_all(gdp$post_gdp, ",","", "")
gdp$post_gdp <- str_replace_all(gdp$post_gdp, ",","", "")

# add country code column
gdp$country_code <- countrycode(gdp$country, origin = "country.name",
destination = 'iso3c')
## Warning in countrycode_convert(sourcevar = sourcevar, origin = origin,
destination = dest, : Some values were not matched unambiguously: Channel
Islands, Created from: World Development Indicators
## Series : GDP (current US$)
##
##
##
## , Kosovo, Turkiye
# add in region
gdp <- gdp %>%
  mutate(country_code = ifelse(country == "Turkiye", "TUR",
                                ifelse(country == "Kosovo", "XKX",
country_code))) %>%
  mutate(country_code = case_when(country_code %in% vax_clean$country_code ~
country_code)) %>%
  filter(!is.na(country_code)) %>%

```

```

inner_join(vax_clean, join_by(country_code))

## add in gdp group
gdp$gdp_group <- cut(as.numeric(gdp$post_gdp),
                      breaks = quantile(as.numeric(gdp$post_gdp),
probs = c(0, 1/3, 2/3, 1), na.rm = TRUE),
                      include.lowest = TRUE,
                      labels = c("Low", "Medium", "High"))
## Warning in cut(as.numeric(gdp$post_gdp), breaks =
## quantile(as.numeric(gdp$post_gdp), : NAs introduced by coercion
## Warning in quantile(as.numeric(gdp$post_gdp), probs = c(0, 1/3, 2/3, 1), :
NAs
## introduced by coercion
gdp_clean <- gdp %>%
  rename(country = country.x) %>%
  select(country, country_code, region, pre_gdp, post_gdp, gdp_group) %>%
  mutate(pre_gdp = as.numeric(pre_gdp), post_gdp = as.numeric(post_gdp))
## Warning: There were 2 warnings in `mutate()` .
## The first warning was:
## i In argument: `pre_gdp = as.numeric(pre_gdp)` .
## Caused by warning:
## ! NAs introduced by coercion
## i Run `dplyr::last_dplyr_warnings()` to see the 1 remaining warning.
rm(gdp)

# View data
head(gdp_clean) %>%
  kable() %>%
  kable_styling() %>%
  scroll_box(width = "100%", height = "400px")

```

country	country_code	region	pre_gdp	post_gdp	gdp_group
Afghanistan	AFG	Asia	14266499430	14266499430	Low
Albania	ALB	Europe	17930565119	17930565119	Low
Algeria	DZA	Africa	163472387987	163472387987	Medium
Angola	AGO	Africa	66505129989	66505129989	Medium
Argentina	ARG	South America	487902572164	487902572164	High
Armenia	ARM	Asia	13878908629	13878908629	Low

Join Data Frames

```

all_data <- gdp_clean %>%
  select(country_code, pre_gdp, post_gdp, gdp_group) %>%
  full_join(vax_clean, join_by(country_code)) %>%
  full_join(life_clean, join_by(country_code))

all_data <- all_data %>%
  select(country.x, region.x, partial, full, vax_group, pre_gdp, post_gdp,
gdp_group, pre_lifeexp, post_lifeexp) %>%
  rename(country = country.x, region = region.x)

## missing data
colSums(is.na(all_data)) %>%

```

```

kable() %>%
  kable_styling()

  x

country      0
region       0
partial       0
full         0
vax_group    0
pre_gdp      5
post_gdp     5
gdp_group   5
pre_lifeexp  1
post_lifeexp 1

## filter out missing data
all_data <- all_data %>%
  filter(!is.na(pre_gdp), !is.na(post_gdp), !is.na(pre_lifeexp), !is.na(post_
lifeexp))

# View data
head(all_data) %>%
  kable() %>%
  kable_styling() %>%
  scroll_box(width = "100%", height = "300px")

country region parti ful vax_gro pre_gdp post_gdp gdp_gro pre_lifee post_lifee
n        al l l up          up           xp           xp

```

country	region	parti ful	vax_gro	pre_gdp	post_gdp	gdp_gro	pre_lifee	post_lifee	
n		al l	l up			up	xp	xp	
Afghanistan	Asia	33	31	Low	142664994 30	142664994 30	Low	63.6	62.0
Albania	Europe	47	45	Medium	179305651 19	179305651 19	Low	79.3	76.5
Algeria	Africa	18	15	Low	163472387 987	163472387 987	Medium	76.5	76.4
Angola	Africa	48	26	Low	665051299 89	665051299 89	Medium	62.4	61.6
Argentina	South America	92	78	High	487902572 164	487902572 164	High	77.3	75.4
Armenia	Asia	38	33	Low	138789086 29	138789086 29	Low	75.4	72.0

Research Question 1

Inspect Data

```

data_1 <- all_data %>%
  select(country, region, vax_group, gdp_group, pre_lifeexp, post_lifeexp)

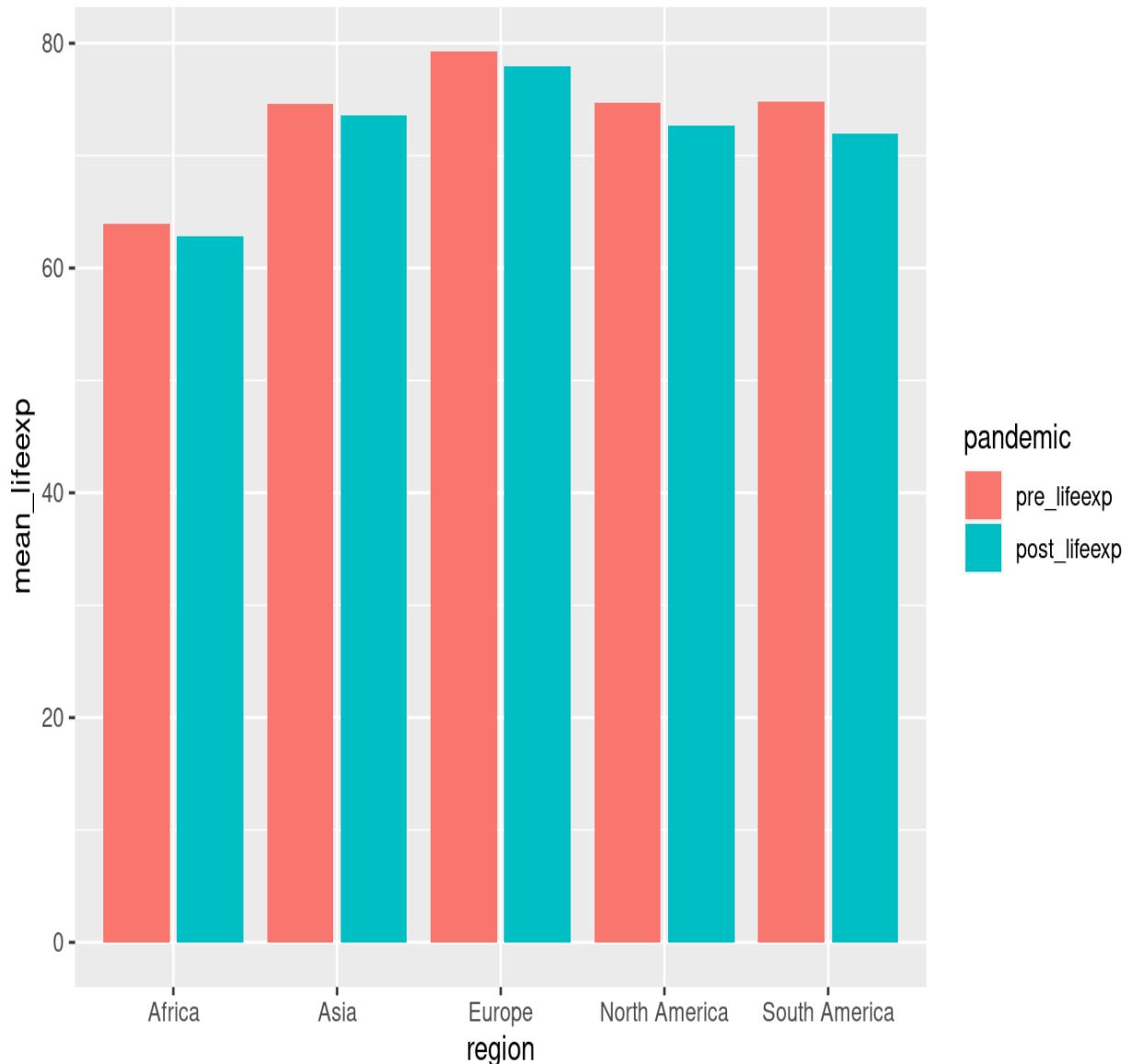
```

```
head(data_1) %>%
  kable() %>%
  kable_styling()

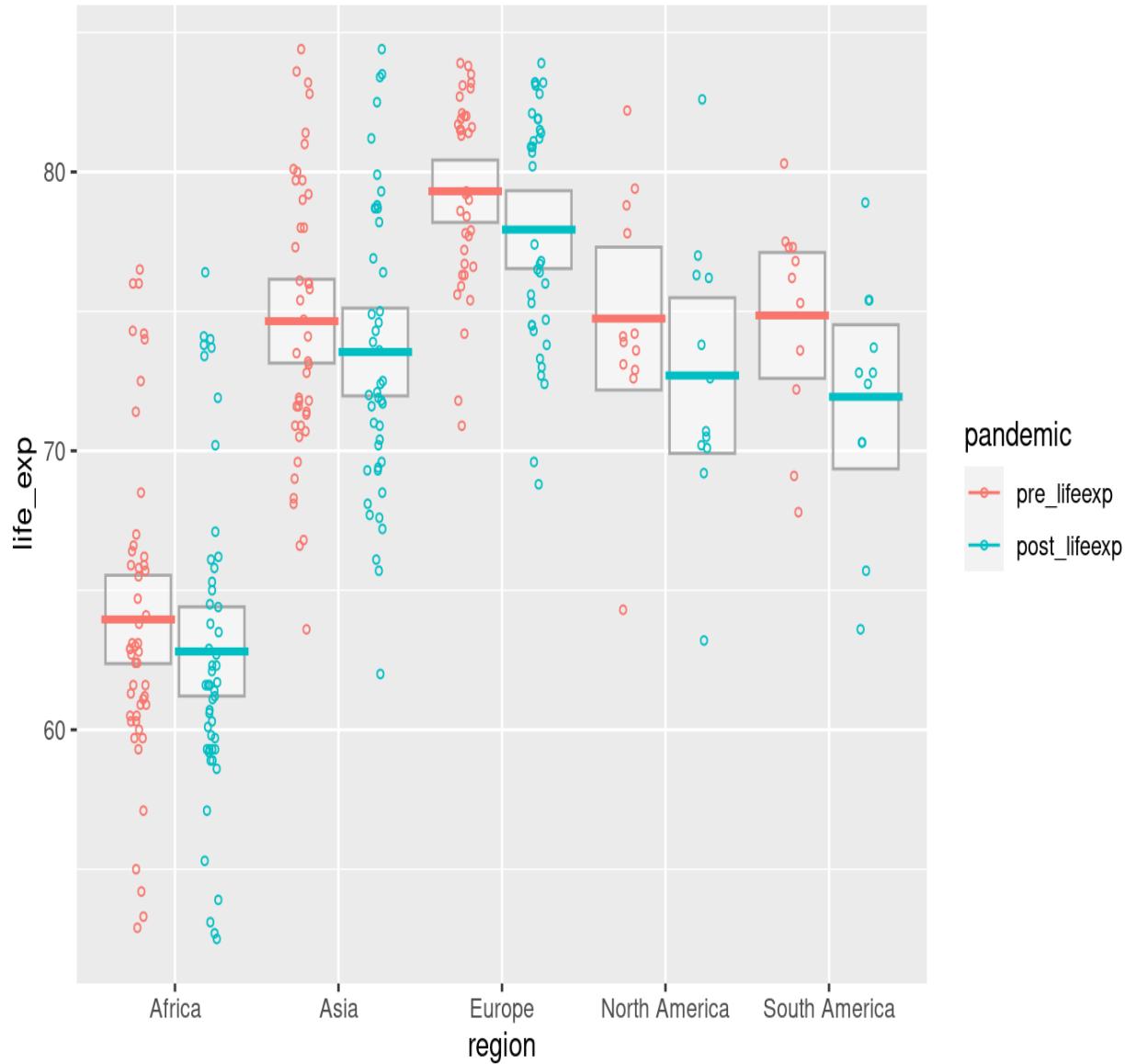
country region vax_group gdp_group pre_lifeexp post_lifeexp
Afghanistan Asia Low Low 63.6 62.0
Albania Europe Medium Low 79.3 76.5
Algeria Africa Low Medium 76.5 76.4
Angola Africa Low Medium 62.4 61.6
Argentina South America High High 77.3 75.4
Armenia Asia Low Low 75.4 72.0
```

Exploratory Visualizations

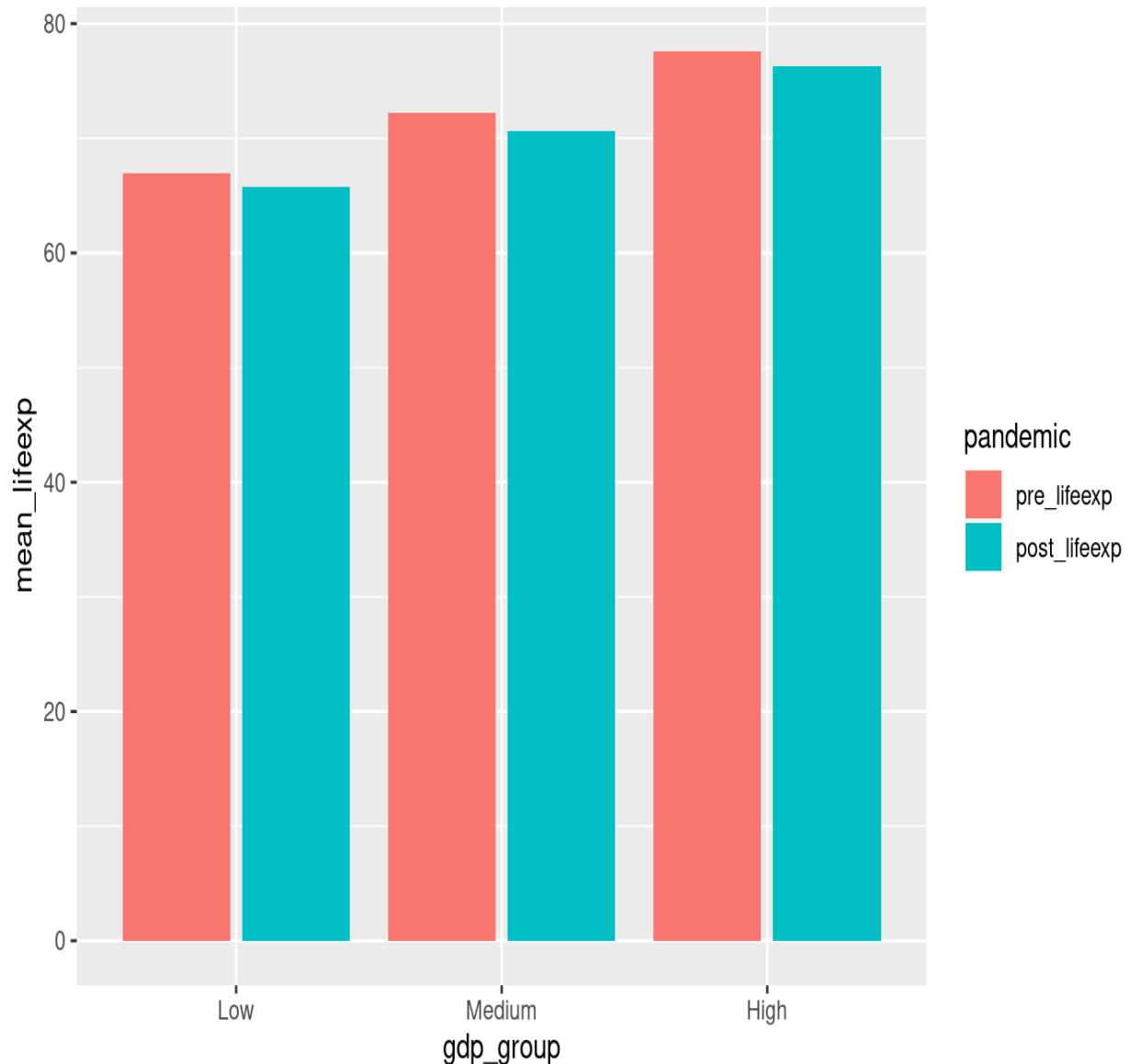
```
# pre vs. post by region
## bar chart
all_data %>%
  select(country, region, pre_lifeexp, post_lifeexp) %>%
  pivot_longer(cols = c(3:4), names_to = "pandemic", values_to =
"life_exp") %>%
  mutate(pandemic = factor(pandemic, levels
=c("pre_lifeexp","post_lifeexp"))) %>%
  group_by(region, pandemic) %>%
  summarize(mean_lifeexp = mean(life_exp)) %>%
  ggplot(aes(x = region, y = mean_lifeexp, fill = pandemic)) +
  geom_bar(stat = "identity", position =position_dodge2(.9))
## `summarise()` has grouped output by 'region'. You can override using the
## `.groups` argument.
```



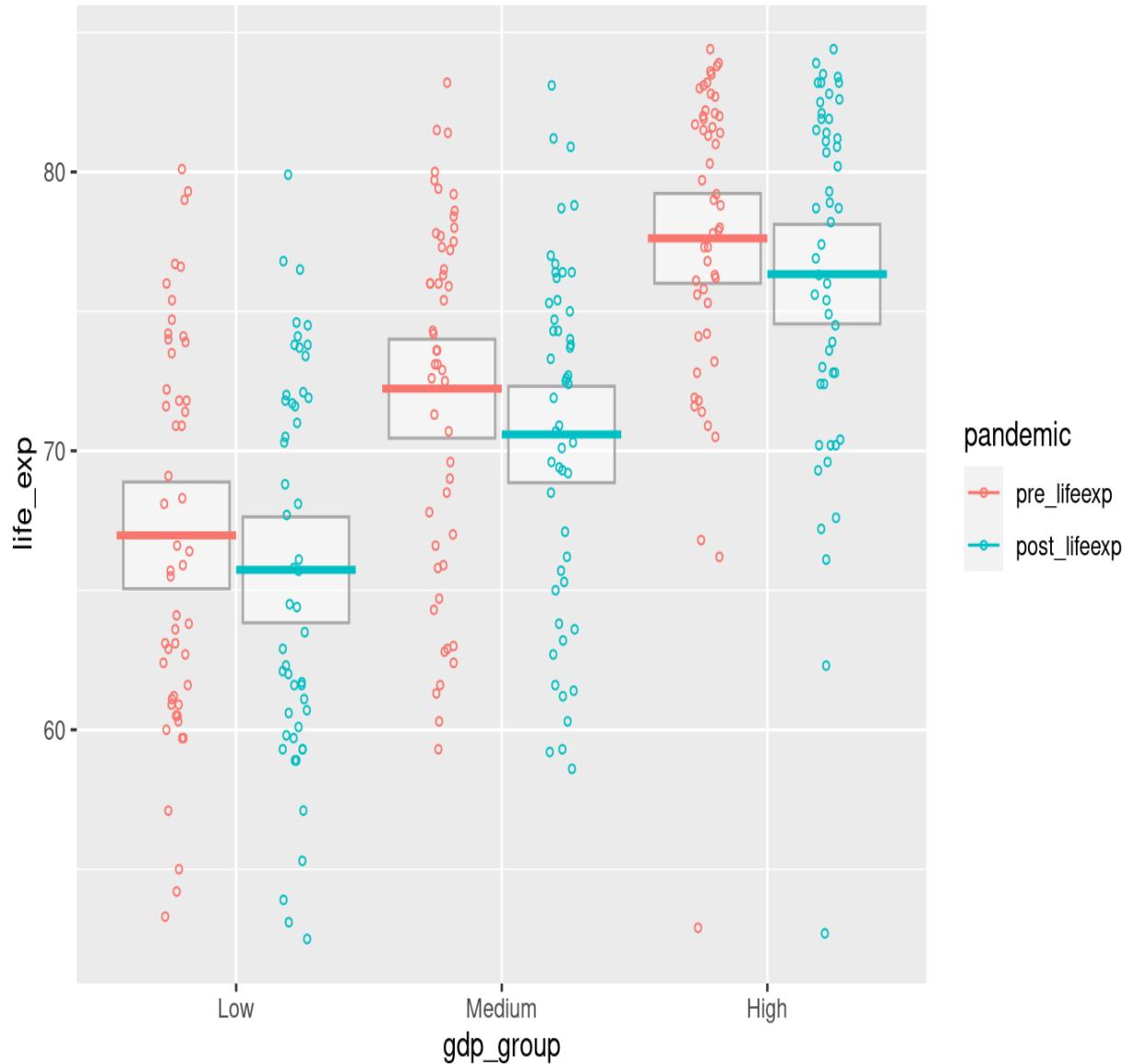
```
## box plot
all_data %>%
  select(country, region, pre_lifeexp, post_lifeexp) %>%
  pivot_longer(cols = c(3:4), names_to = "pandemic", values_to =
"life_exp") %>%
  mutate(pandemic = factor(pandemic, levels =
c("pre_lifeexp", "post_lifeexp"))) %>%
  group_by(region, pandemic) %>%
  ggplot(aes(x = region, y = life_exp, color = pandemic)) +
  geom_pirate(violins = FALSE, bars = FALSE, show.legend = TRUE)
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



```
# post vs. pre by GDP
## bar chart
all_data %>%
  select(country, gdp_group, pre_lifeexp, post_lifeexp) %>%
  pivot_longer(cols = c(3:4), names_to = "pandemic", values_to =
"life_exp") %>%
  mutate(pandemic = factor(pandemic, levels
=c("pre_lifeexp","post_lifeexp")))%>%
  group_by(gdp_group, pandemic) %>%
  summarize(mean_lifeexp = mean(life_exp)) %>%
  ggplot(aes(x = gdp_group, y = mean_lifeexp, fill = pandemic)) +
  geom_bar(stat = "identity", position =position_dodge2(.9))
## `summarise()` has grouped output by 'gdp_group'. You can override using
the
## `.` argument.
```



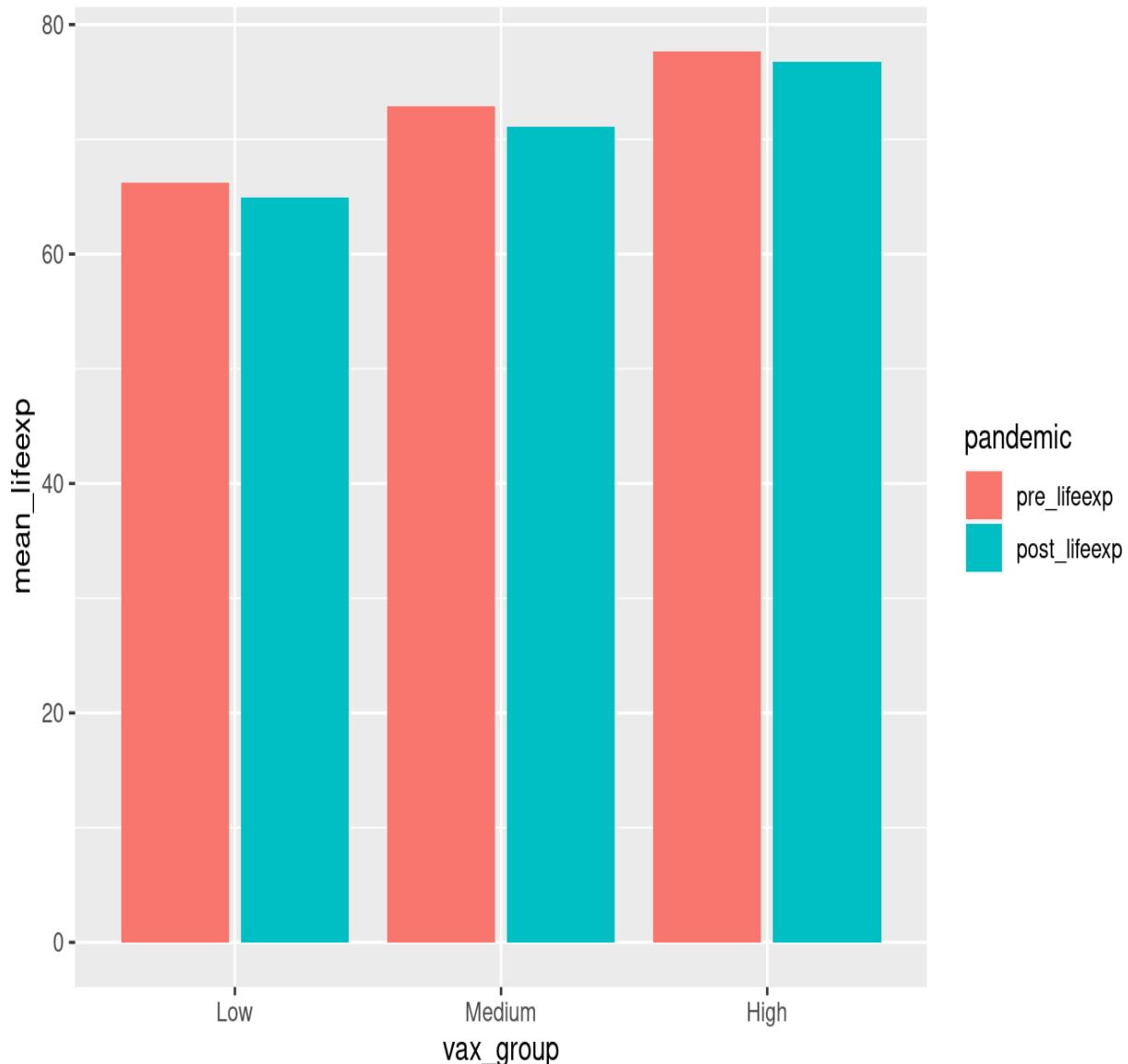
```
## box plot
all_data %>%
  select(country, gdp_group, pre_lifeexp, post_lifeexp) %>%
  pivot_longer(cols = c(3:4), names_to = "pandemic", values_to =
"life_exp") %>%
  mutate(pandemic = factor(pandemic, levels =
c("pre_lifeexp", "post_lifeexp"))) %>%
  group_by(gdp_group, pandemic) %>%
  ggplot(aes(x = gdp_group, y = life_exp, color = pandemic)) +
  geom_pirate(violins = FALSE, bars = FALSE, show.legend = TRUE)
```



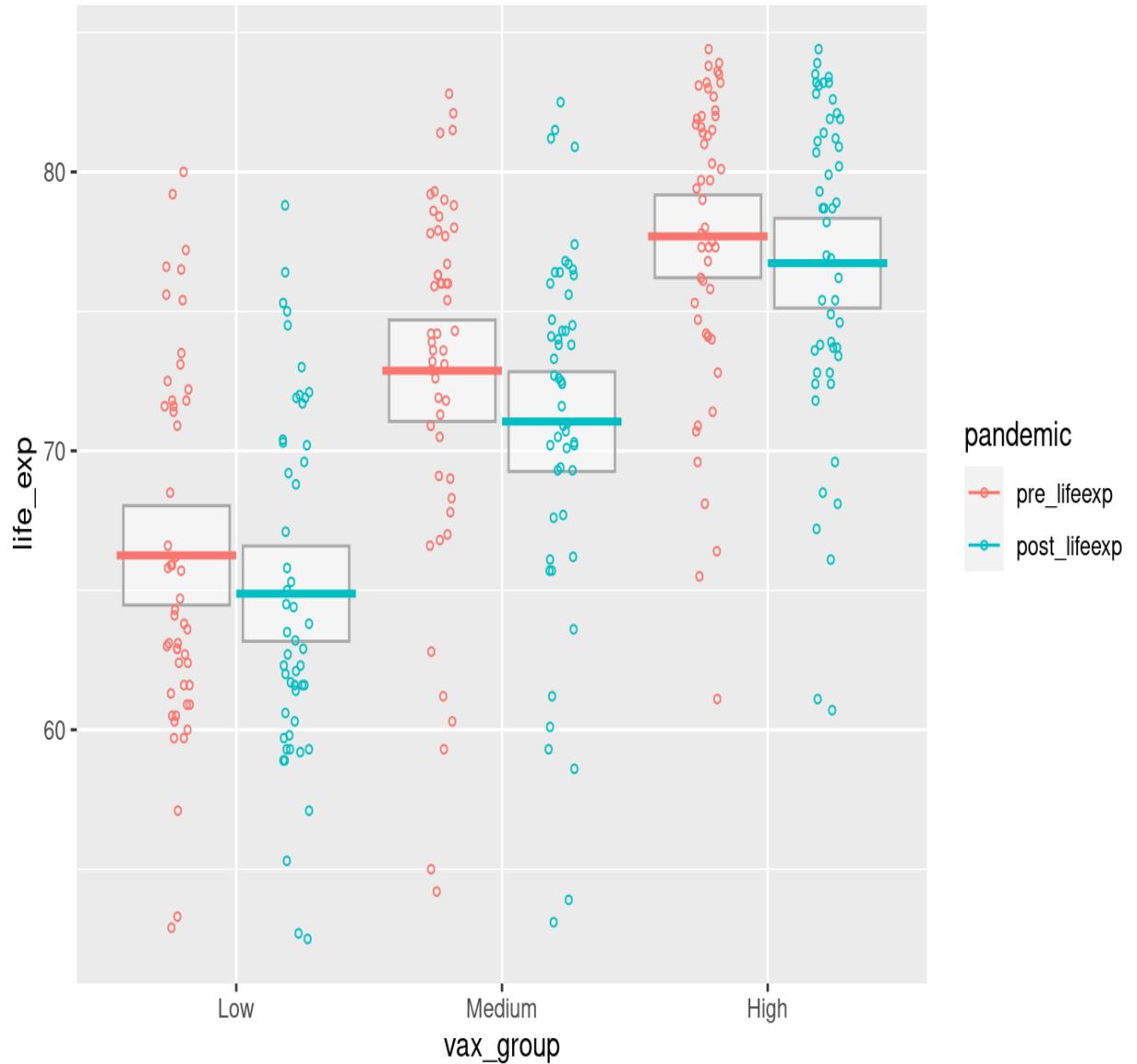
```

# post vs. pre by vax group
## bar chart
all_data %>%
  select(country, vax_group, pre_lifeexp, post_lifeexp) %>%
  pivot_longer(cols = c(3:4), names_to = "pandemic", values_to =
"life_exp") %>%
  mutate(pandemic = factor(pandemic, levels
=c("pre_lifeexp","post_lifeexp")))%>%
  group_by(vax_group, pandemic) %>%
  summarize(mean_lifeexp = mean(life_exp)) %>%
  ggplot(aes(x = vax_group, y = mean_lifeexp, fill = pandemic)) +
  geom_bar(stat = "identity", position =position_dodge2(.9))
## `summarise()` has grouped output by 'vax_group'. You can override using
the
## `.` argument.

```



```
## box plot
all_data %>%
  select(country, vax_group, pre_lifeexp, post_lifeexp) %>%
  pivot_longer(cols = c(3:4), names_to = "pandemic", values_to =
"life_exp") %>%
  mutate(pandemic = factor(pandemic, levels =
c("pre_lifeexp", "post_lifeexp"))) %>%
  group_by(vax_group, pandemic) %>%
  ggplot(aes(x = vax_group, y = life_exp, color = pandemic)) +
  geom_pirate(violins = FALSE, bars = FALSE, show.legend = TRUE)
```



Statistical Analysis

```
# need long format
all_data_long <- all_data %>%
  pivot_longer(cols = c(9,10), names_to = "pandemic", values_to =
"life exp") %>%
  mutate(pandemic = factor(pandemic, levels
=c("pre_lifeexp","post_lifeexp")))
```

Check Assumptions

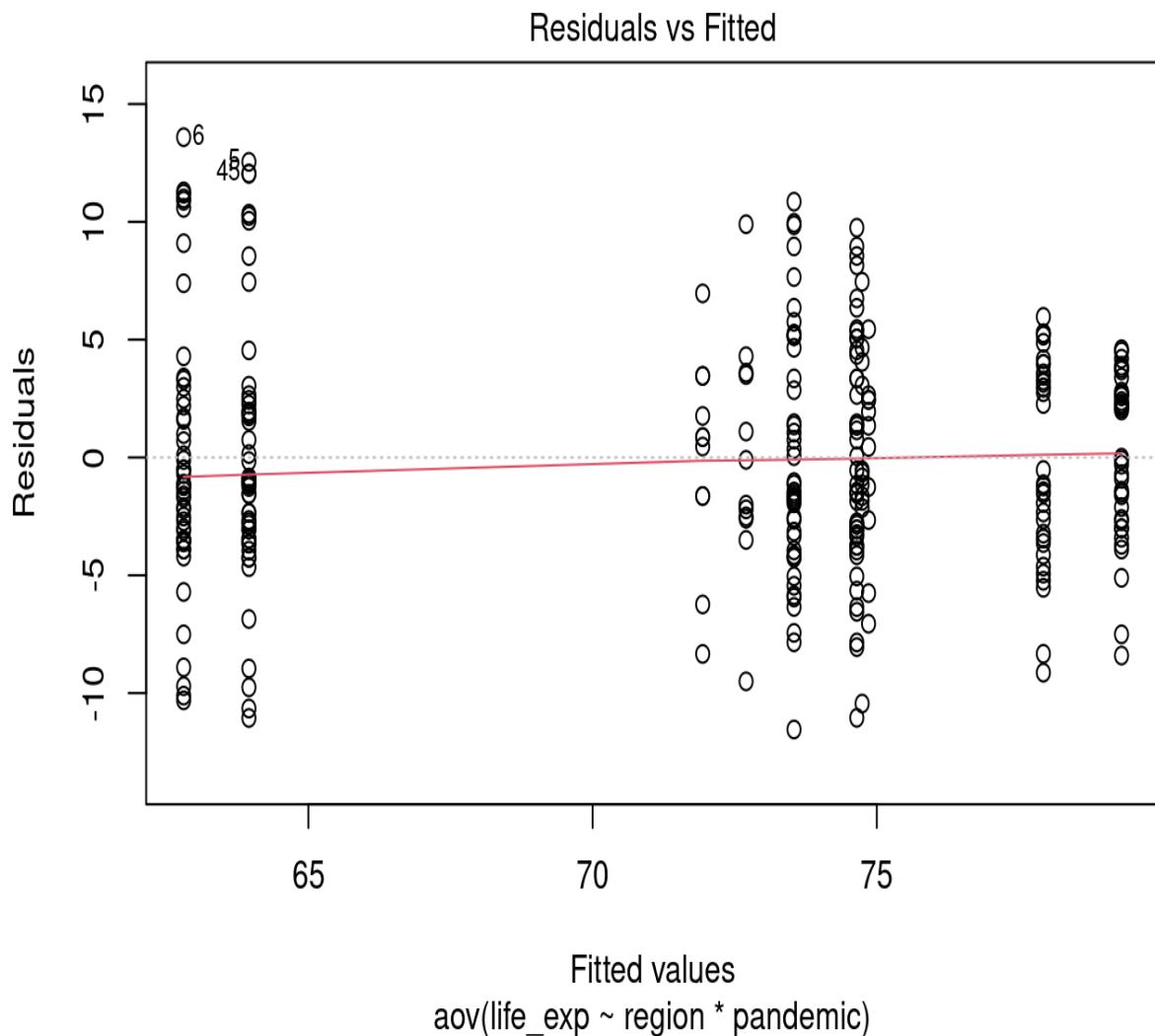
```
# Region
anova_region <- aov(life_exp ~ region * pandemic, data = all_data_long)

## homogeneity of variances
```

```

leveneTest(anova_region)
## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value Pr(>F)
## group     9  0.8932 0.5316
## 296
plot(anova_region,1)

```

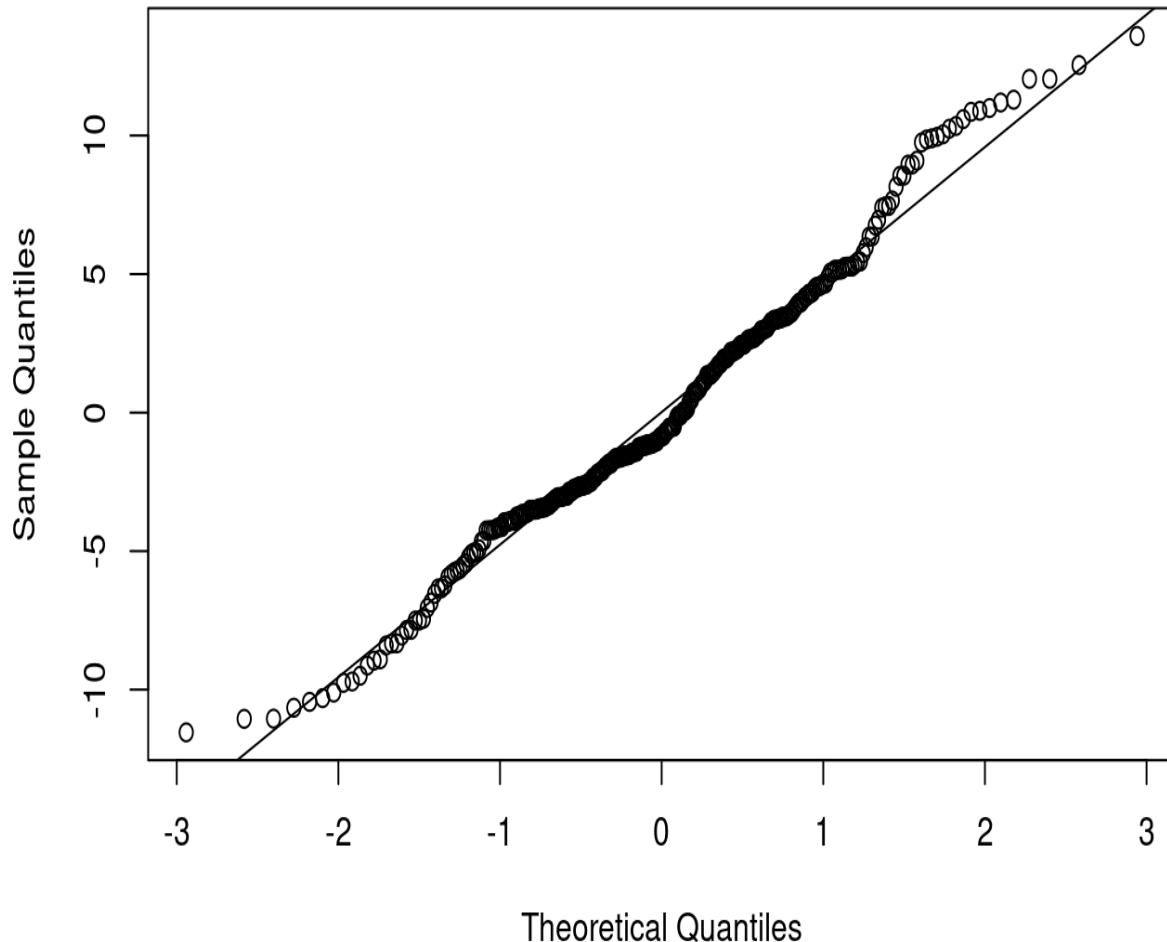


```

## normality of residuals
qqnorm(residuals(anova_region))
qqline(residuals(anova_region))

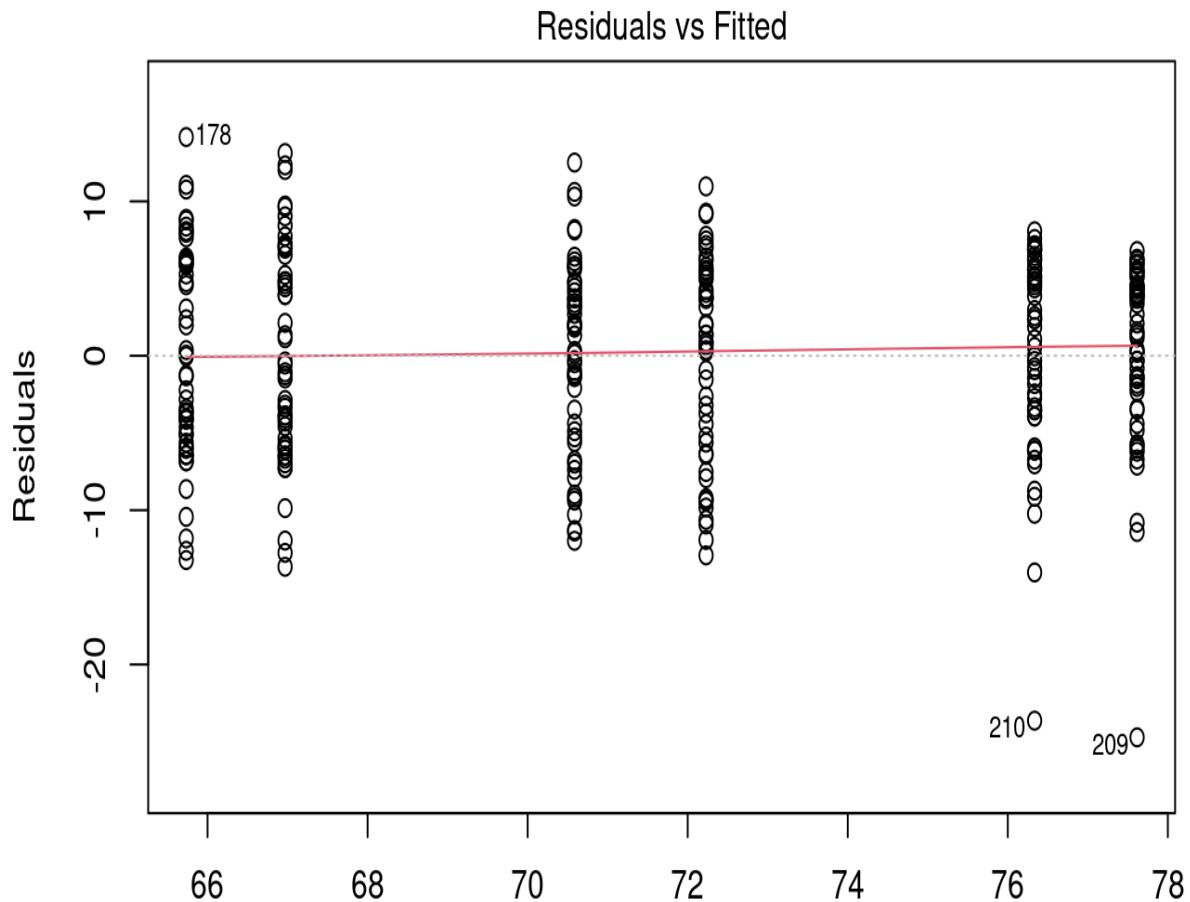
```

Normal Q-Q Plot



```
# GDP Group
anova_gdp <- aov(life_exp ~ gdp_group * pandemic, data = all_data_long)

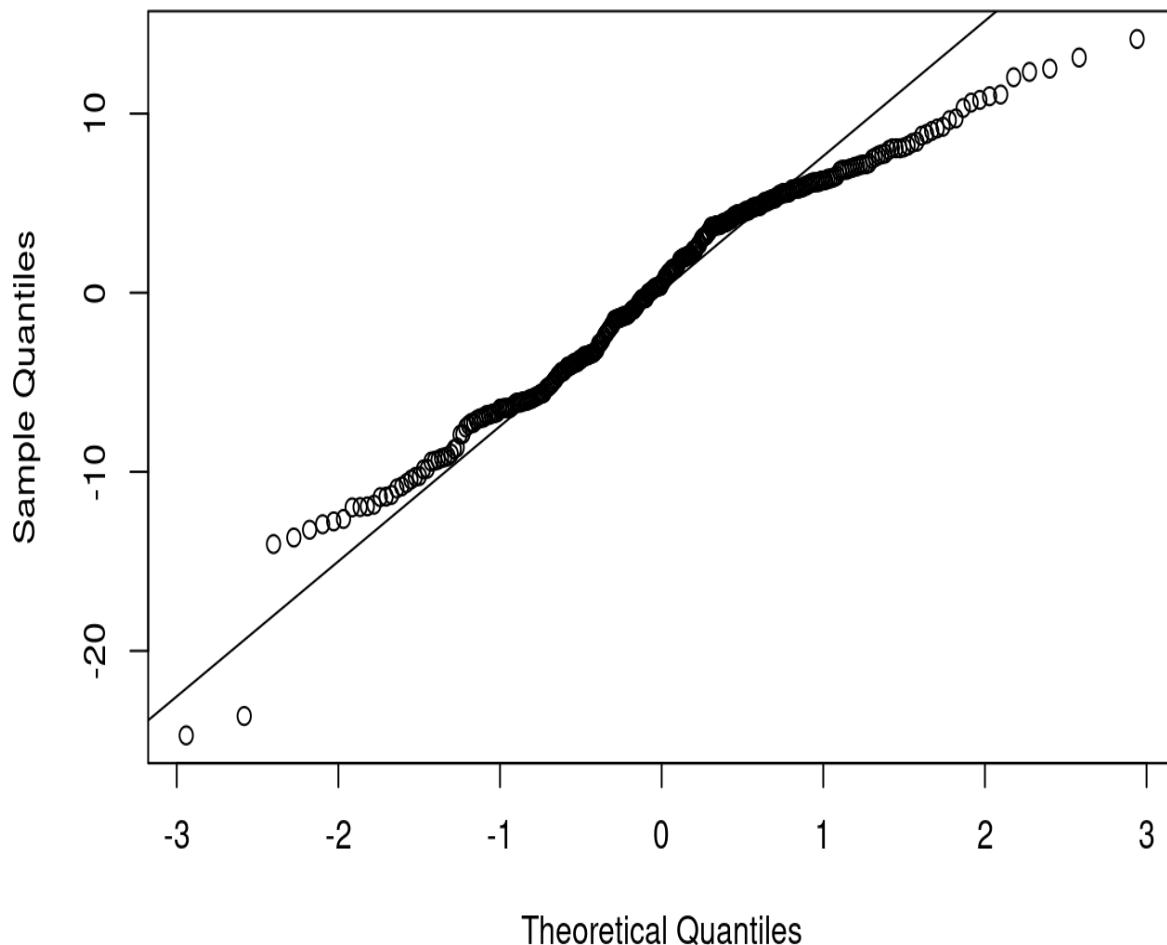
## homogeneity of variances
leveneTest(anova_gdp)
## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value Pr(>F)
## group      5  1.2518 0.2849
##             300
plot(anova_gdp, 1)
```



Fitted values
aov(life_exp ~ gdp_group * pandemic)

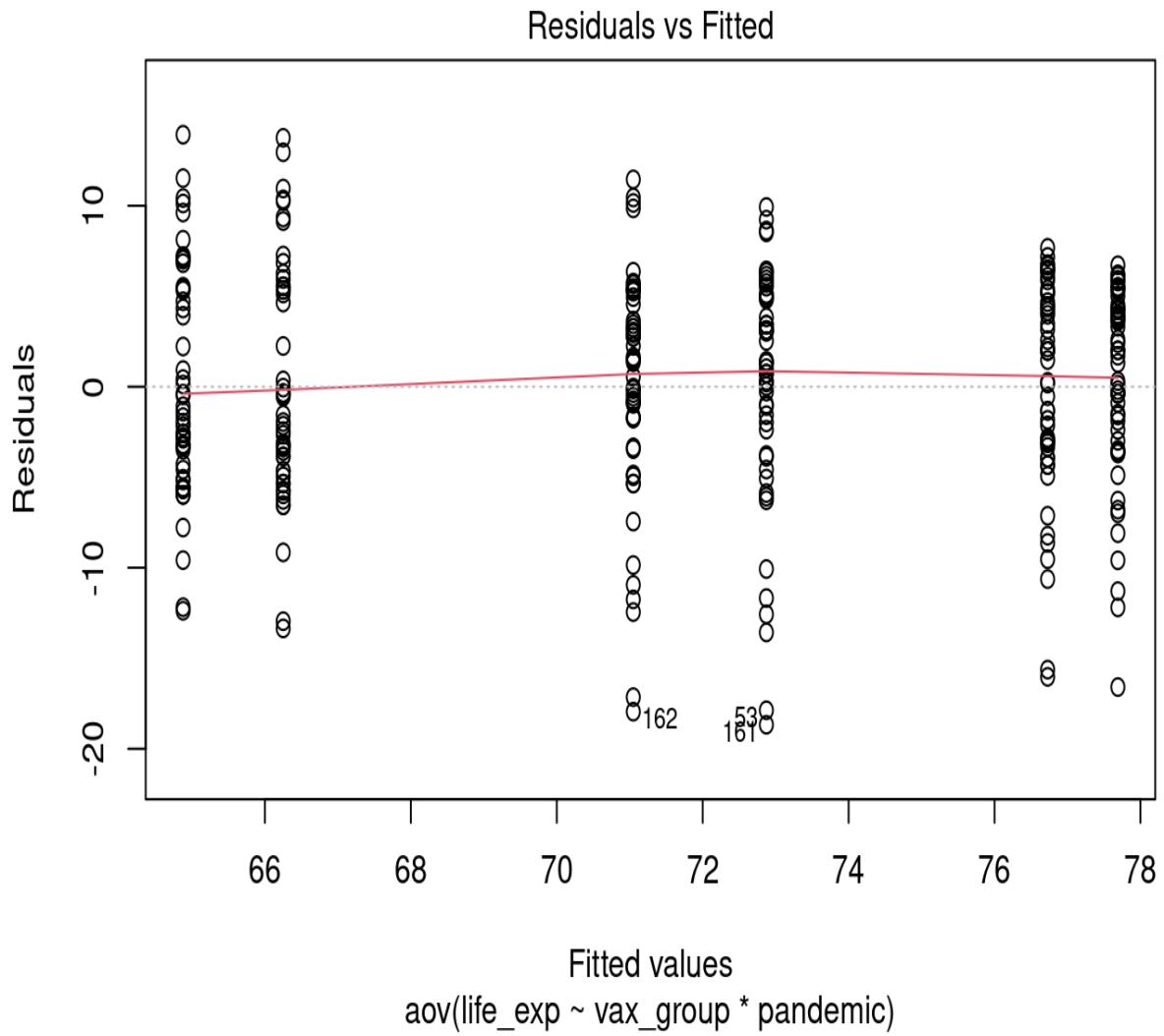
```
## normality of residuals
qqnorm(residuals(anova_gdp))
qqline(residuals(anova_gdp))
```

Normal Q-Q Plot



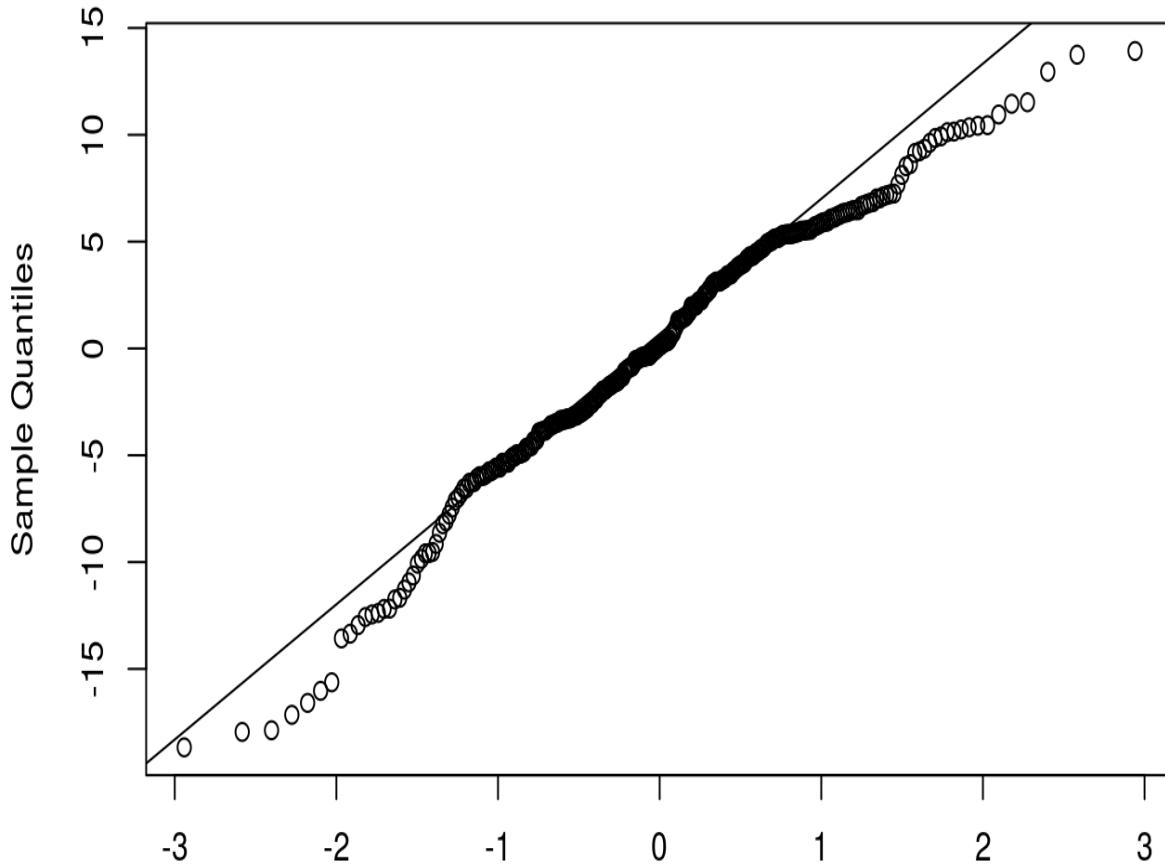
```
# vax Group
anova_vax <- aov(life_exp ~ vax_group * pandemic, data = all_data_long)

## homogeneity of variances
leveneTest(anova_vax)
## Levene's Test for Homogeneity of Variance (center = median)
##          Df F value Pr(>F)
## group      5  0.3068 0.9087
##             300
plot(anova_vax,1)
```



```
## normality of residuals
qqnorm(residuals(anova_vax))
qqline(residuals(anova_vax))
```

Normal Q-Q Plot



Theoretical Quantiles

ANOVA

```
# two way anova
anova <- aov(life_exp ~ (region + gdp_group + vax_group) * pandemic, data = all_data_long)

## results
summary(anova)
##                                     Df Sum Sq Mean Sq F value    Pr(>F)
## region                           4 11030  2757.4 147.948 < 2e-16 ***
## gdp_group                         2    853   426.3  22.875 6.02e-10 ***
## vax_group                          2   1206   602.9  32.348 2.12e-13 ***
## pandemic                          1    147   147.0   7.888  0.00532 **
## region:pandemic                  4     19    4.7   0.250  0.90942
## gdp_group:pandemic                2      2    1.0   0.052  0.94930
## vax_group:pandemic                2     12    5.9   0.316  0.72901
## Residuals                         288  5368   18.6
```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
# significant diff in life exp across region
# significant diff in life exp across gdp groups
#significant difference in life exp across vax groups
# pandemic significantly decreased life exp
# No significant interaction effects -- effect of pandemic on life exp did
not vary across regions or gdp groups or vax groups

```

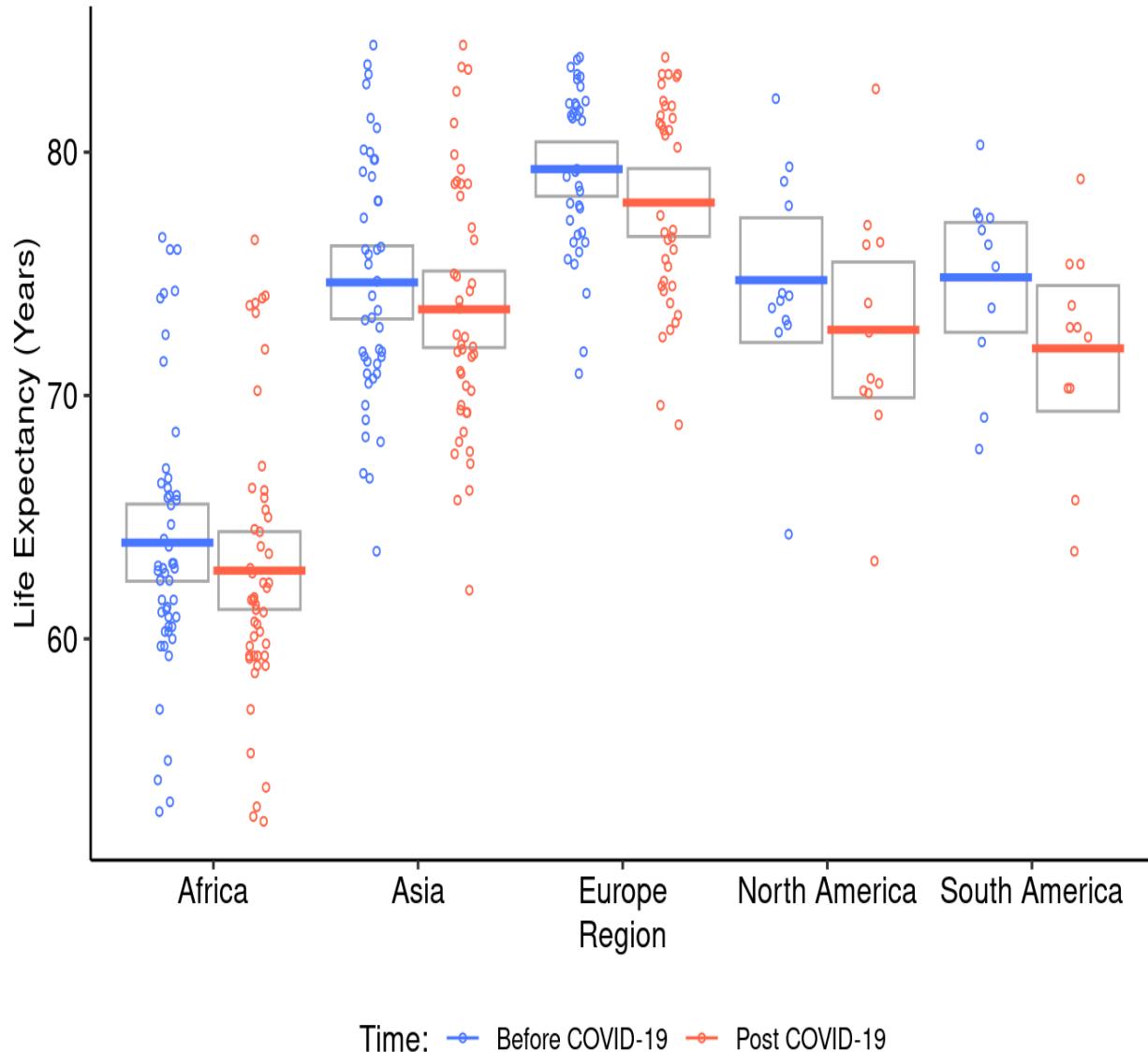
Final Visualizations of Results

Region

```

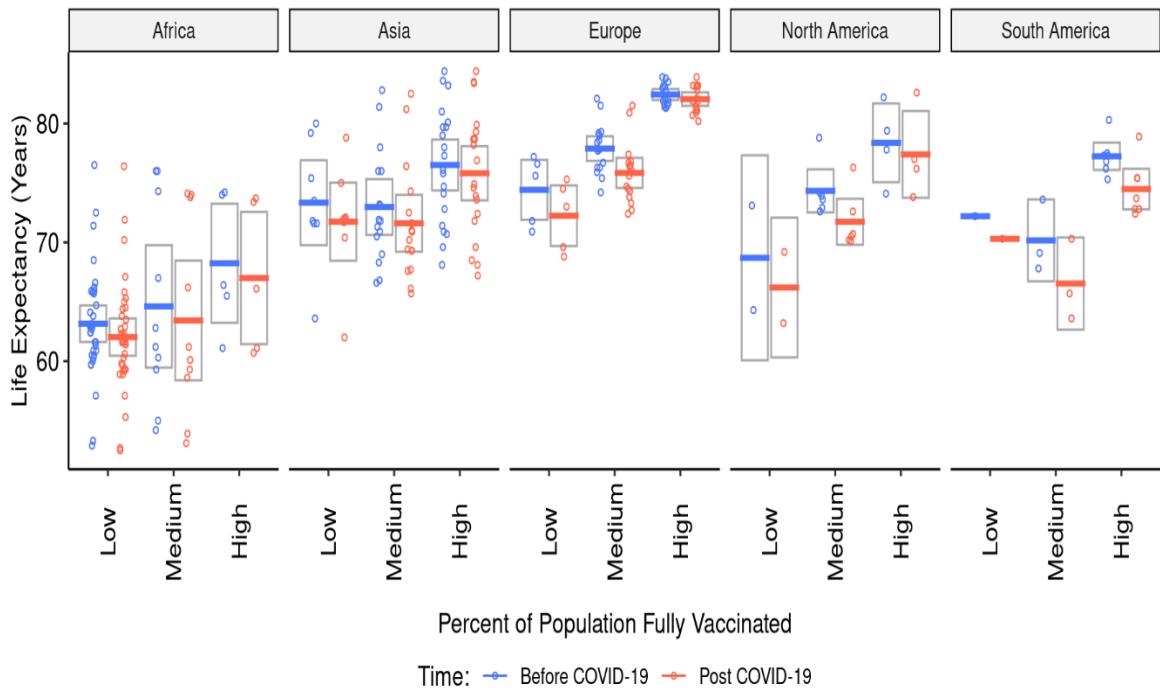
all_data_long %>%
  ggplot(aes(x = region, y = life_exp, color = pandemic)) +
  geom_pirate(violins = FALSE, bars = FALSE, show.legend = TRUE) +
  theme_pubr() +
  labs( y = "Life Expectancy (Years)", x = "Region") +
  scale_color_manual(name = "Time:", labels = c("Before COVID-19", "Post
COVID-19"),
                     values = c("royalblue1", "tomato")) +
  theme(legend.position = "bottom")

```



Vax Group by Region

```
all_data_long %>%
  ggplot(aes(x = vax_group, y = life_exp, color = pandemic)) +
  geom_pirate(violins = FALSE, bars = FALSE, show.legend = TRUE) +
  theme_pubr() +
  labs(y = "Life Expectancy (Years)", x = "Percent of Population Fully Vaccinated") +
  scale_color_manual(name = "Time:", labels = c("Before COVID-19", "Post COVID-19"),
                     values = c("royalblue1", "tomato")) +
  theme(legend.position = "bottom") +
  facet_grid(~region) +
  theme(axis.text.x = element_text(angle = 90, hjust = 0.5, vjust = 0.5),
        plot.margin = margin(1,1,1,1, "cm"), axis.title.x = element_text(vjust = -3))
## Warning: Removed 2 rows containing missing values (`geom_tile()`).
```



Research Question 2

Inspect Data

```
data_2 <- all_data %>%
  select(country, region, full, post_gdp, post_lifeexp) %>%
  rename(gdp = post_gdp)

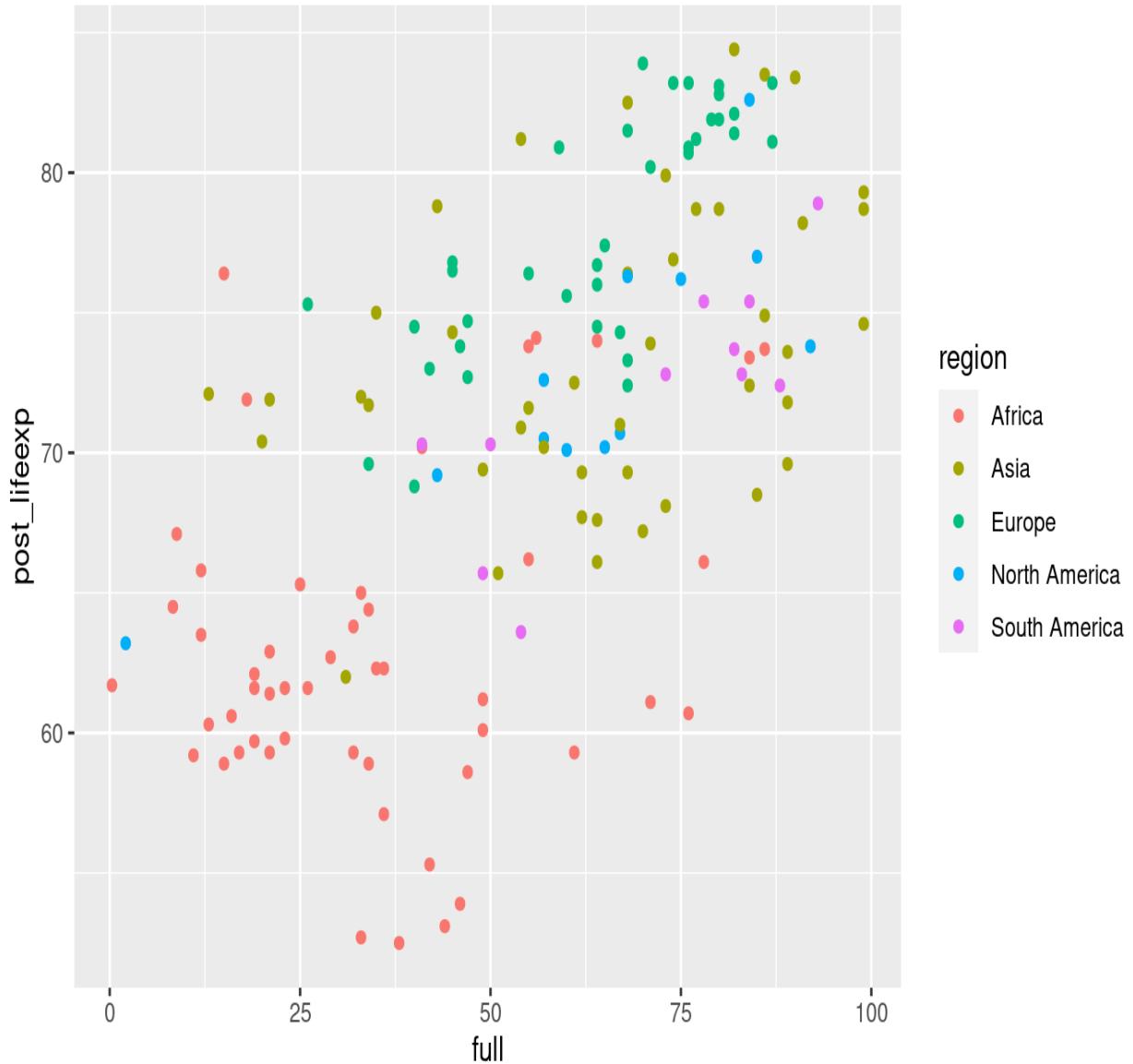
head(data_2) %>%
  kable() %>%
  kable_styling()

country region      full      gdp post_lifeexp
Afghanistan Asia        31 14266499430      62.0
Albania     Europe       45 17930565119      76.5
Algeria     Africa        15 163472387987    76.4
Angola      Africa       26 66505129989      61.6
Argentina   South America 78 487902572164    75.4
Armenia     Asia         33 13878908629      72.0
```

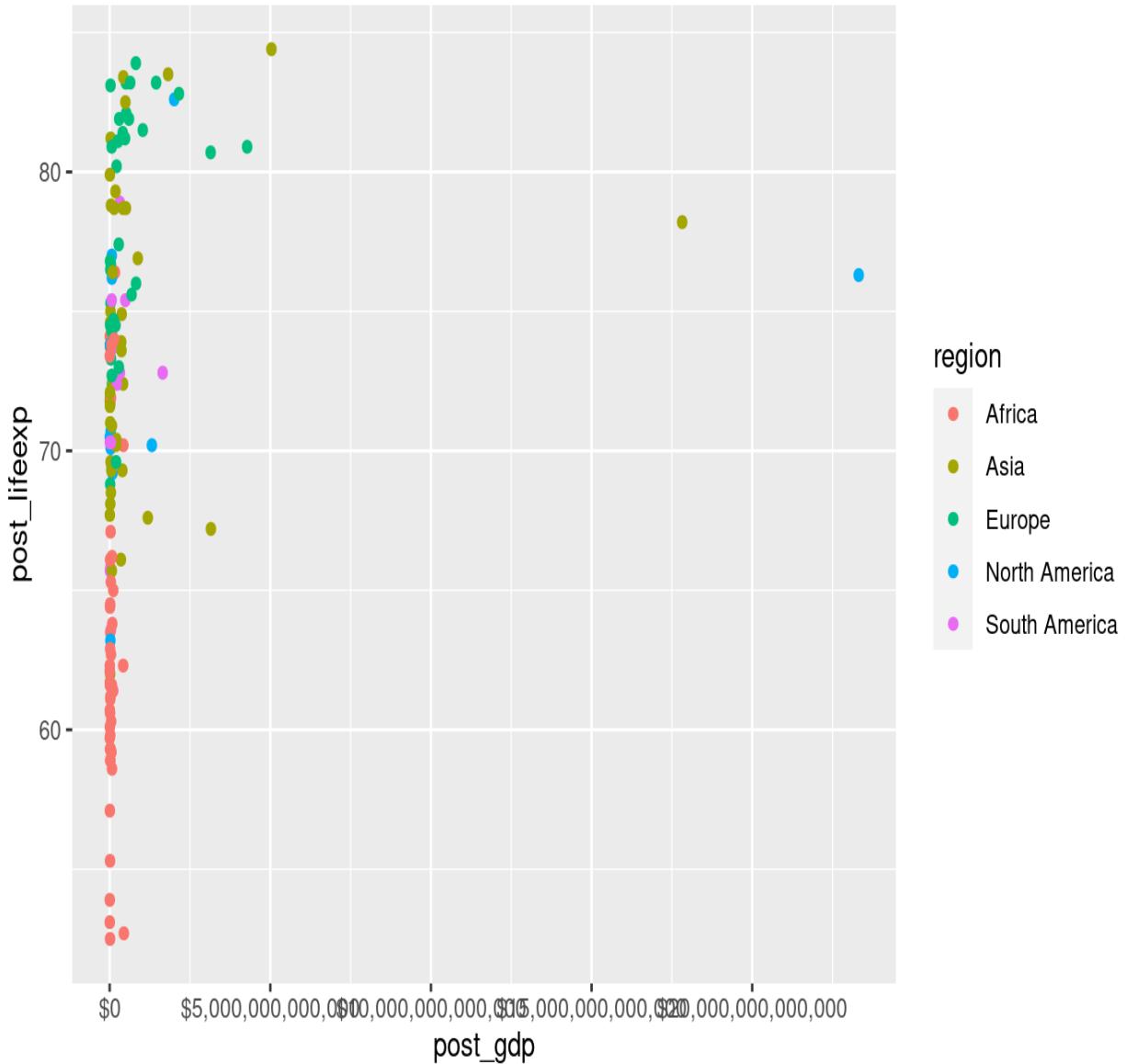
Exploratory Visualizations

```
# scatterplot vax vs. life
all_data %>%
```

```
ggplot(aes(x = full, y = post_lifeexp, color = region)) +  
  geom_point()
```

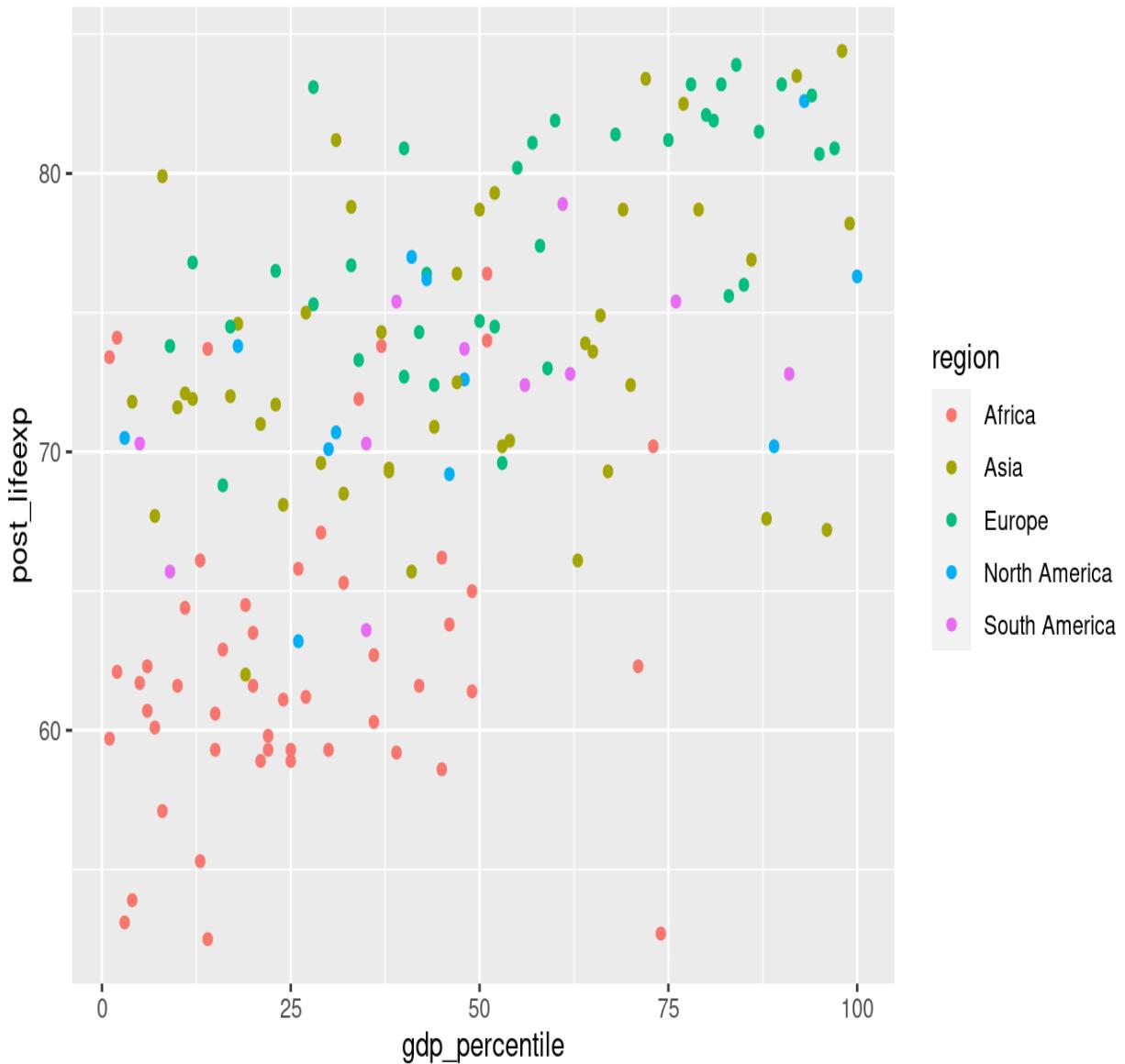


```
# scatterplot gdp vs. life  
all_data%>%  
  ggplot(aes(x = post_gdp, y = post_lifeexp, color = region)) +  
  geom_point() + scale_x_continuous(labels = dollar)
```

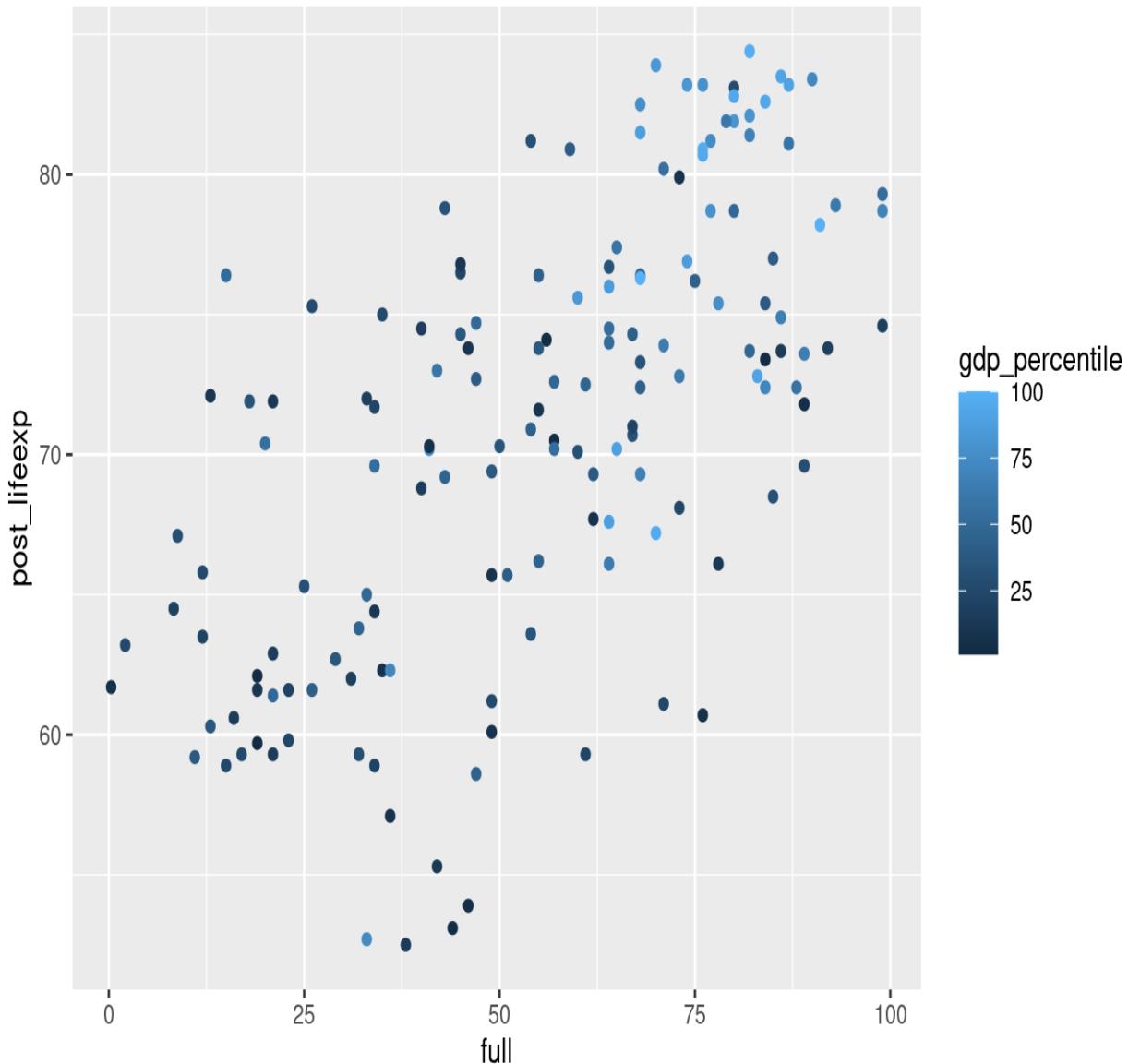


```
## obvi GDP in USD isn't good -- change to percentile
data_2$gdp_percentile <- ntile(data_2$gdp, 100)

# scatterplot gdp percentile vs. life
data_2 %>%
  ggplot(aes(x = gdp_percentile, y = post_lifeexp, color = region)) +
  geom_point()
```



```
# scatterplot life exp vs. life exp with color by gdp percentile
data_2 %>%
  ggplot(aes(y = post_lifeexp, x = full, color = gdp_percentile)) +
  geom_point() +
  theme(legend.position = "right")
```



Statistical Analysis - Regression & Correlation

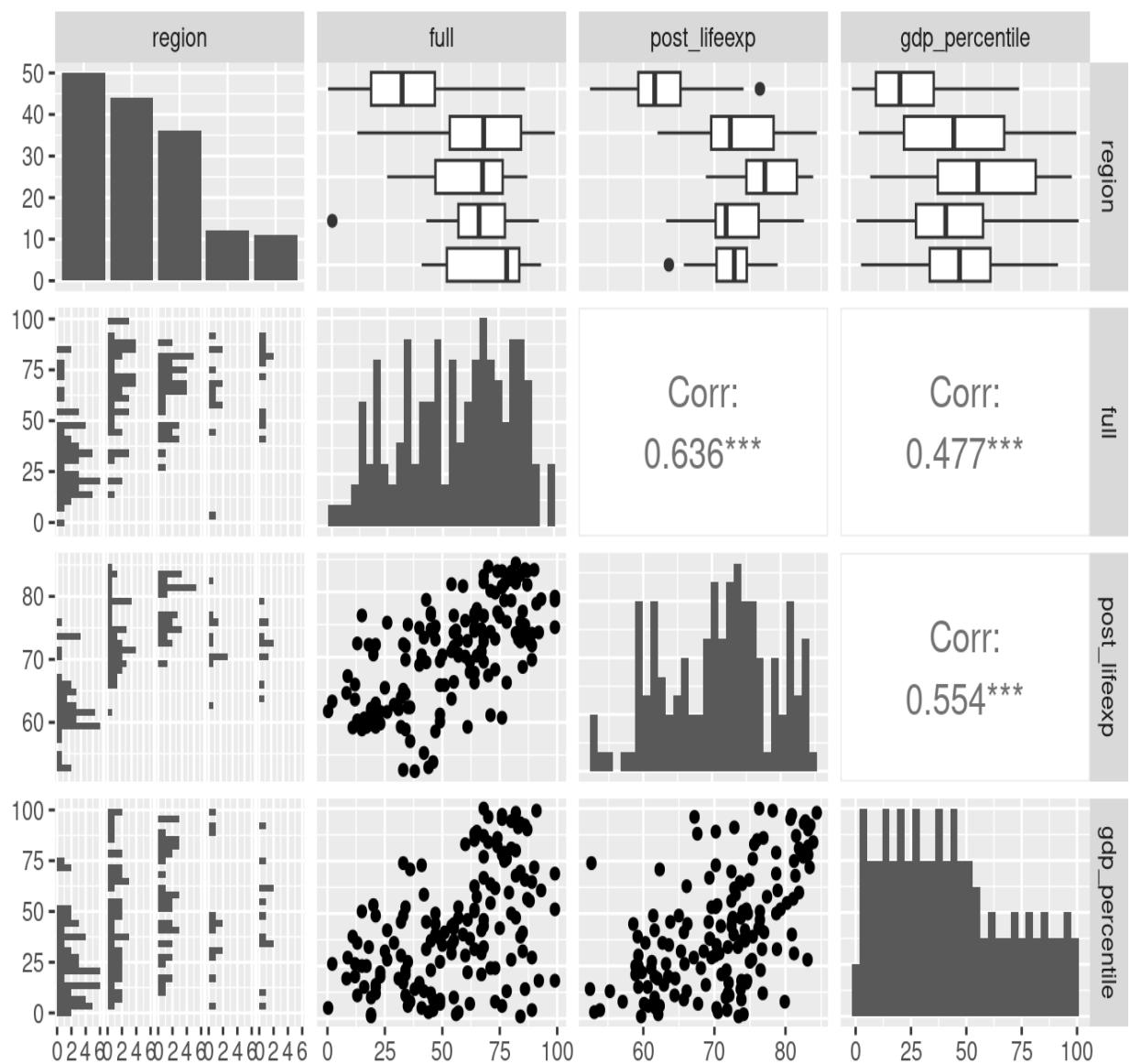
Correlation

```
library(GGally)
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
data_2 %>%
select(-country, - gdp )%>%
ggpairs( upper = list(continuous = wrap("cor", size = 5)), lower =
list(continuous = "points"), diag = list(continuous = "barDiag"))
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

```



```

## significant relationship between life exp (2021) and % of pop fully vax (r = 0.636 -- strong, positive)
## significant relationship between gdp percentile and % of pop fully vax (r = 0.477 -- strong, positive)
## significant relationship between life exp (2021) and gdp percentile (r = 0.554 -- strong, positive)

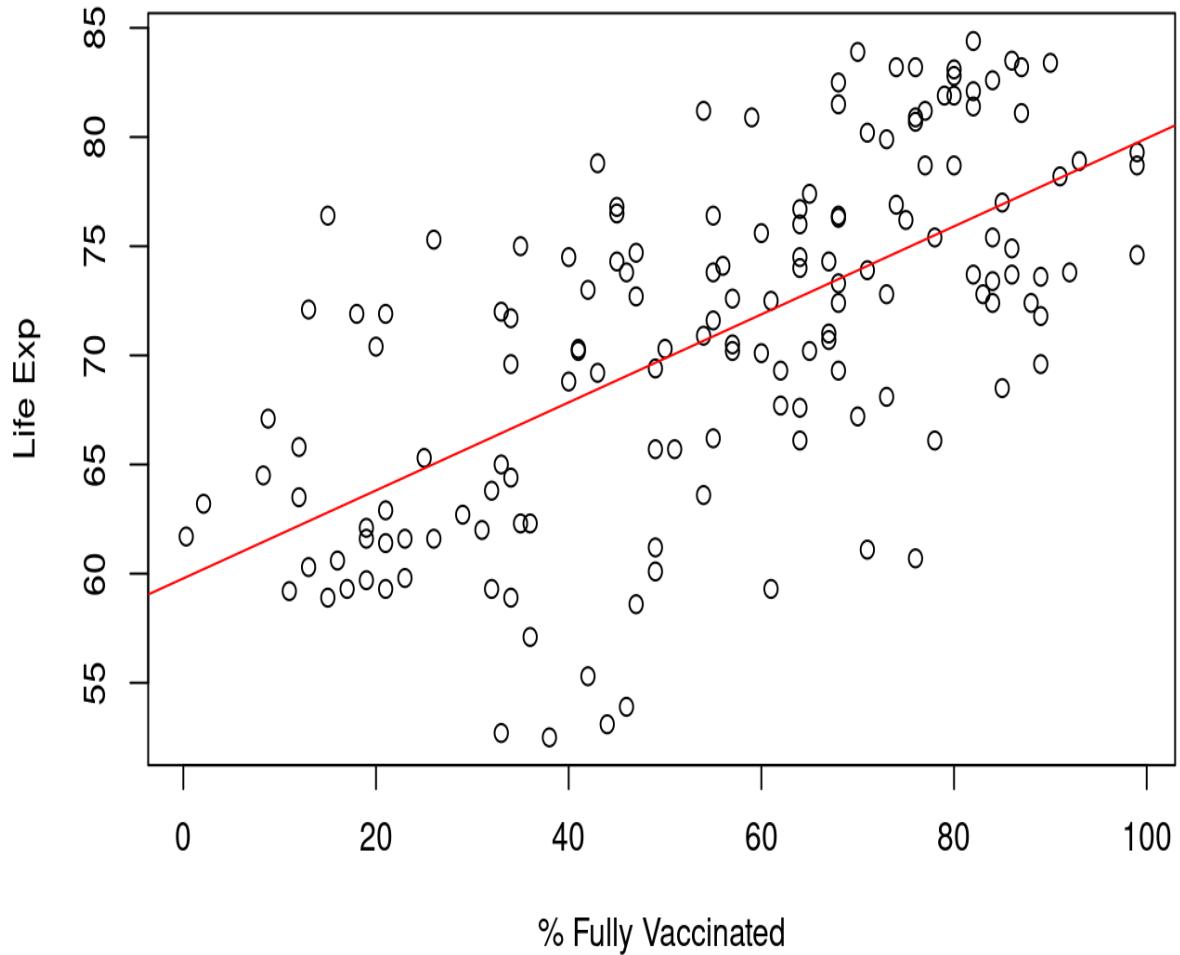
```

Regression

```
# how well vaccination coverage predicts life exp (2021)
```

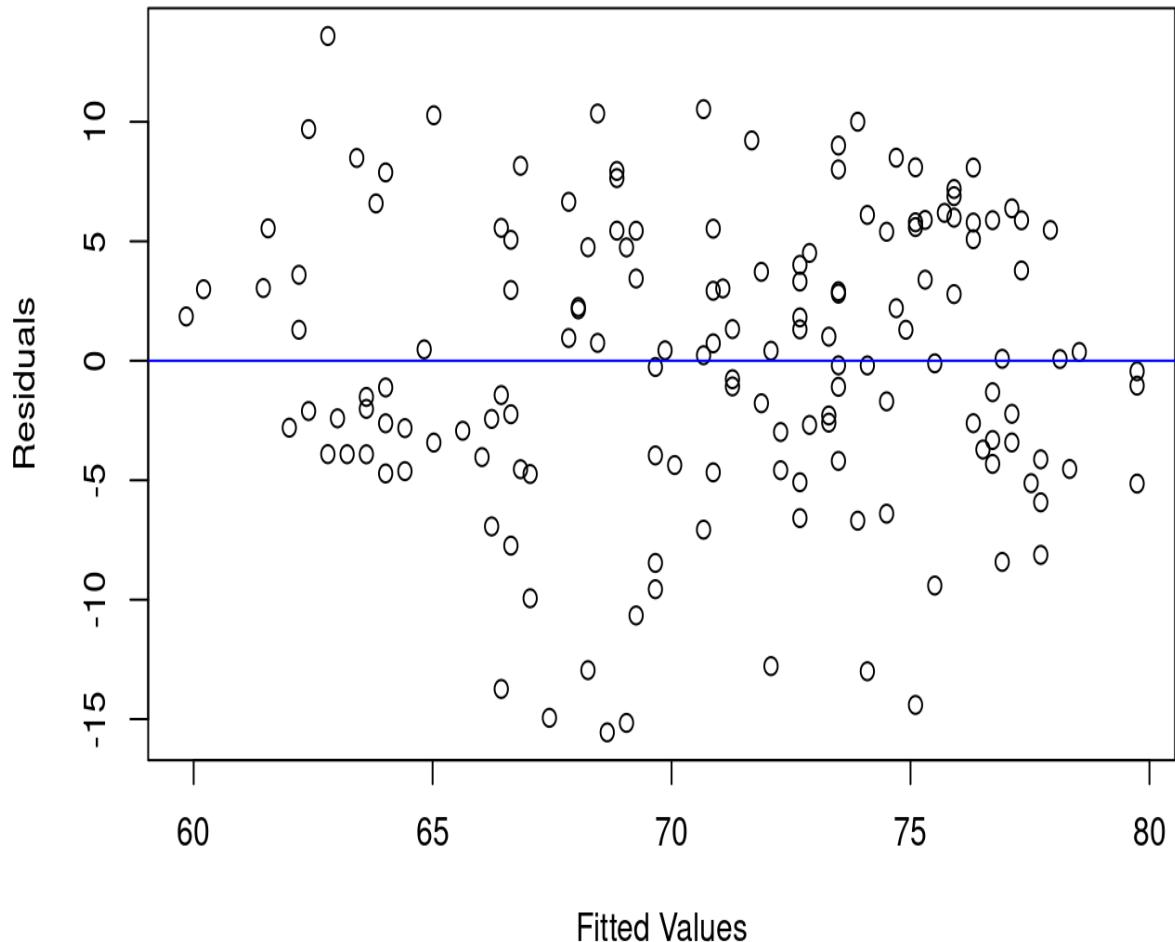
```
modell <- lm(post_lifeexp ~ full, data = data_2)
summary(modell)
##
## Call:
## lm(formula = post_lifeexp ~ full, data = data_2)
##
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -15.5537 -3.9140  0.0813  5.0860 13.5923 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 59.78390   1.20126  49.77 <2e-16 ***
## full        0.20159   0.01991  10.13 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.076 on 151 degrees of freedom
## Multiple R-squared:  0.4044, Adjusted R-squared:  0.4005 
## F-statistic: 102.5 on 1 and 151 DF,  p-value: < 2.2e-16
## asses model fit & assumptions
plot(data_2$full, data_2$post_lifeexp, main="Life Expectancy (years) vs.
Fully Vaccinated", xlab="% Fully Vaccinated", ylab="Life Exp")
abline(modell, col="red")
```

Life Expectancy (years) vs. Fully Vaccinated



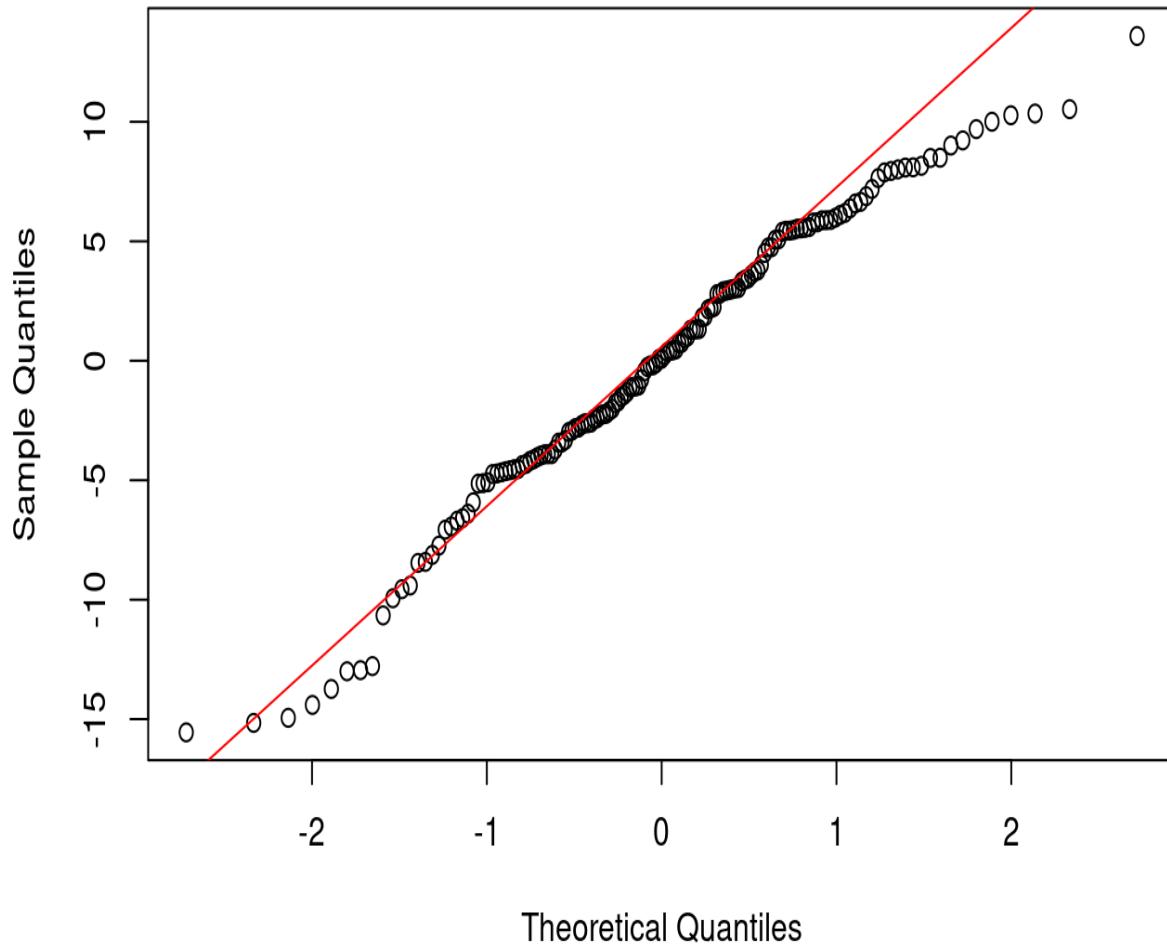
```
## Residuals vs Fitted (to check for homoscedasticity and linearity)
plot(modell$fitted.values, residuals(modell), xlab="Fitted Values",
ylab="Residuals", main="Residuals vs Fitted")
abline(h=0, col="blue")
```

Residuals vs Fitted



```
### Normal Q-Q Plot (to check for normality of residuals)
qqnorm(residuals(model1))
qqline(residuals(model1), col="red")
```

Normal Q-Q Plot



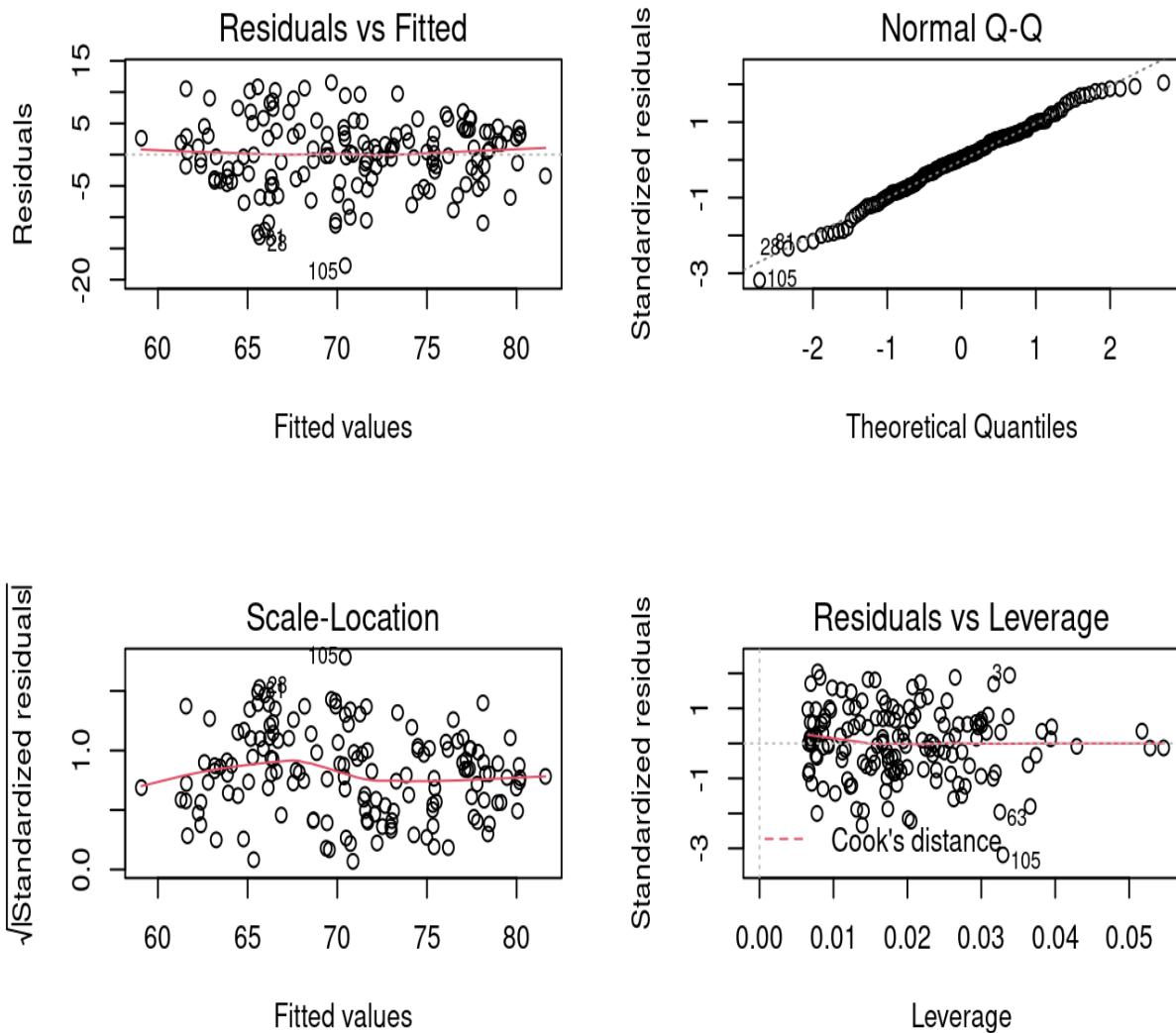
Theoretical Quantiles

```
# model with gdp too
model2 <- lm(post_lifeexp ~ full + gdp_percentile, data=data_2)
summary(model2)
##
## Call:
## lm(formula = post_lifeexp ~ full + gdp_percentile, data = data_2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -17.7418  -3.8244   0.2686   3.6379  11.5282 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 58.56452   1.14805 51.012 < 2e-16 ***
## full        0.15262    0.02112  7.225 2.36e-11 ***
## gdp_percentile 0.09244    0.01902  4.859 2.94e-06 ***
```

```

## ---
## Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.666 on 150 degrees of freedom
## Multiple R-squared:  0.4855, Adjusted R-squared:  0.4786
## F-statistic: 70.76 on 2 and 150 DF,  p-value: < 2.2e-16
par(mfrow=c(2,2))
plot(model2)

```



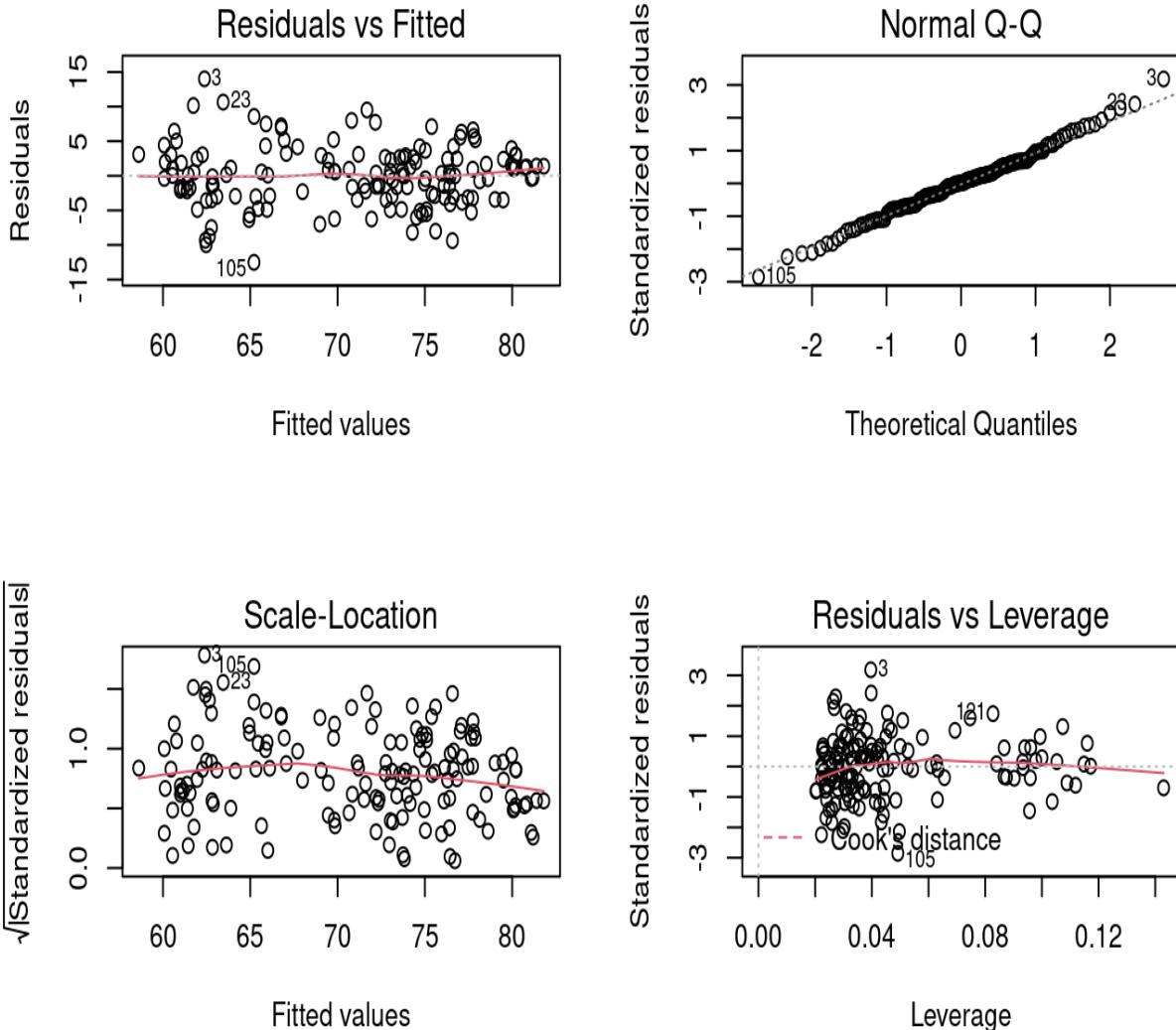
```

# model with region and gdp percentile
model3 <- lm(post_lifeexp ~ full + region+ gdp_percentile, data=data_2)
summary(model3)
##
## Call:
## lm(formula = post_lifeexp ~ full + region + gdp_percentile, data = data_2)
## 
```

```

## Residuals:
##      Min       1Q   Median      3Q     Max
## -12.5122 -2.9487  0.0517  2.6474 14.0248
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)           58.31877   0.91224  63.929 < 2e-16 ***
## full                  0.08969   0.01907   4.702 5.91e-06 ***
## regionAsia            6.87410   1.08823   6.317 3.04e-09 ***
## regionEurope          10.90376   1.15191   9.466 < 2e-16 ***
## regionNorth America  6.22157   1.54017   4.040 8.62e-05 ***
## regionSouth America  4.80036   1.63313   2.939 0.003824 **
## gdp_percentile       0.05316   0.01565   3.396 0.000881 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.495 on 146 degrees of freedom
## Multiple R-squared:  0.6848, Adjusted R-squared:  0.6718
## F-statistic: 52.86 on 6 and 146 DF,  p-value: < 2.2e-16
par(mfrow=c(2,2))
plot(model3)

```



```
## Compare models:
models <- list(model1, model2, model3)
modelsummary(models, stars = TRUE)
```

(Intercept)	
full	
gdp_percentile	
regionAsia	

Model 1	Model 2	Model 3
59.784***	58.565***	58.319***
(1.201)	(1.148)	(0.912)
0.202***	0.153***	0.090***
(0.020)	(0.021)	(0.019)
0.092***	0.053***	
(0.019)	(0.016)	
	6.874***	
	(1.088)	

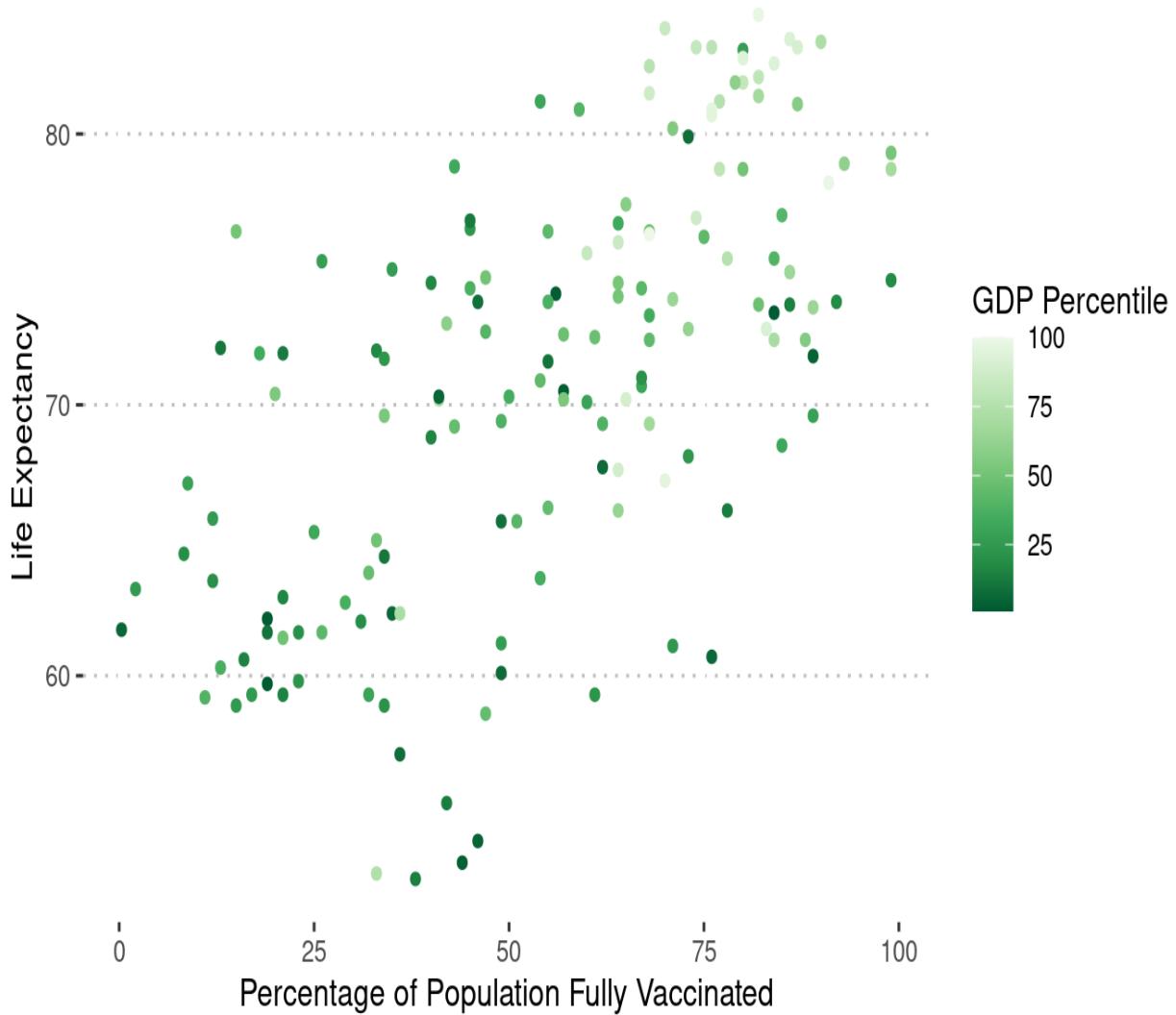
	Model 1	Model 2	Model 3
regionEurope			10.904*** (1.152)
regionNorth America			6.222*** (1.540)
regionSouth America			4.800** (1.633)
Num.Obs.	153	153	153
R2	0.404	0.485	0.685
R2 Adj.	0.401	0.479	0.672
AIC	990.3	969.9	903.0
BIC	999.4	982.1	927.2
Log.Lik.	-492.153	-480.969	-443.480
F	102.547	70.759	52.865
RMSE	6.04	5.61	4.39

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Viz of Results

```
##  
ggplot(data_2, aes(x = full, y = post_lifeexp, color = gdp_percentile)) +  
  geom_point() +  
  labs(title = "Life Expectancy vs Full Vaccination Coverage by GDP  
Percentile",  
       x = "Percentage of Population Fully Vaccinated",  
       y = "Life Expectancy",  
       color = "GDP Percentile") +  
  theme_pubclean() + scale_color_distiller(palette = "Greens") +  
  theme(legend.position = "right")
```

Life Expectancy vs Full Vaccination Coverage by GDP Percentile



-
1. ### World Bank. (2024). World Development Indicators. Retrieved from <https://databank.worldbank.org/source/world-development-indicators>.