

# Assignment 7

## Assignment 7 (alternate)

```
library(tidyverse)
```

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

```
## v ggplot2 3.3.5    v purrr  0.3.4
## v tibble  3.1.5    v dplyr  1.0.7
## v tidyr   1.1.4    v stringr 1.4.0
## v readr   2.0.2    v forcats 0.5.1
```

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

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

```
library(plotly)
```

```
##
```

```
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
```

```
##
```

```
##    last_plot
```

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

```
##
```

```
##    filter
```

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

```
##
```

```
##    layout
```

```
library(rvest)
```

```
##
```

```
## Attaching package: 'rvest'
```

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

```
##
```

```
##    guess_encoding
```

For this assignment, I want you to use the Census ACS API to download and plot data. Complete the following steps:

1. Scrape the data on GDP per capita by country from wikipedia here: [https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_GDP\\_\(PPP\)\\_per\\_capita](https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP)_per_capita) Turn this data into a data frame.

```
gdp_wiki<-"https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP)_per_capita"

gdp_percap<-read_html(gdp_wiki)%>%html_table()

gdp_percap<-gdp_percap[[2]]

gdp_percap<-
  gdp_percap%>%
  select(1,8)%>%
  rename(country=1,gdp_percap=2)%>%
  slice(-1)%>%
  mutate(gdp_percap=parse_number(gdp_percap))%>%
  mutate(country=str_remove(country, fixed("(more)")))%>%
  mutate(country=str_trim(country))
```

```
## Warning: 1 parsing failure.
## row col expected actual
## 226  -- a number      N/A
```

2. Download the data on tertiary education in 2019 (first table) by country from wikipedia here: [https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_tertiary\\_education\\_attainment](https://en.wikipedia.org/wiki/List_of_countries_by_tertiary_education_attainment)

Turn this data into a data frame.

```
educ_wiki<- "https://en.wikipedia.org/wiki/List_of_countries_by_tertiary_education_attainment"

educ<-read_html(educ_wiki)%>%html_table()

educ<-educ[[1]]

educ<-educ%>%
  select(1,2)%>%
  rename(country=1,educ_level=2)%>%
  slice(-1)%>%
  mutate(educ_level=parse_number(educ_level))
```

3. Join the two data frames you created.

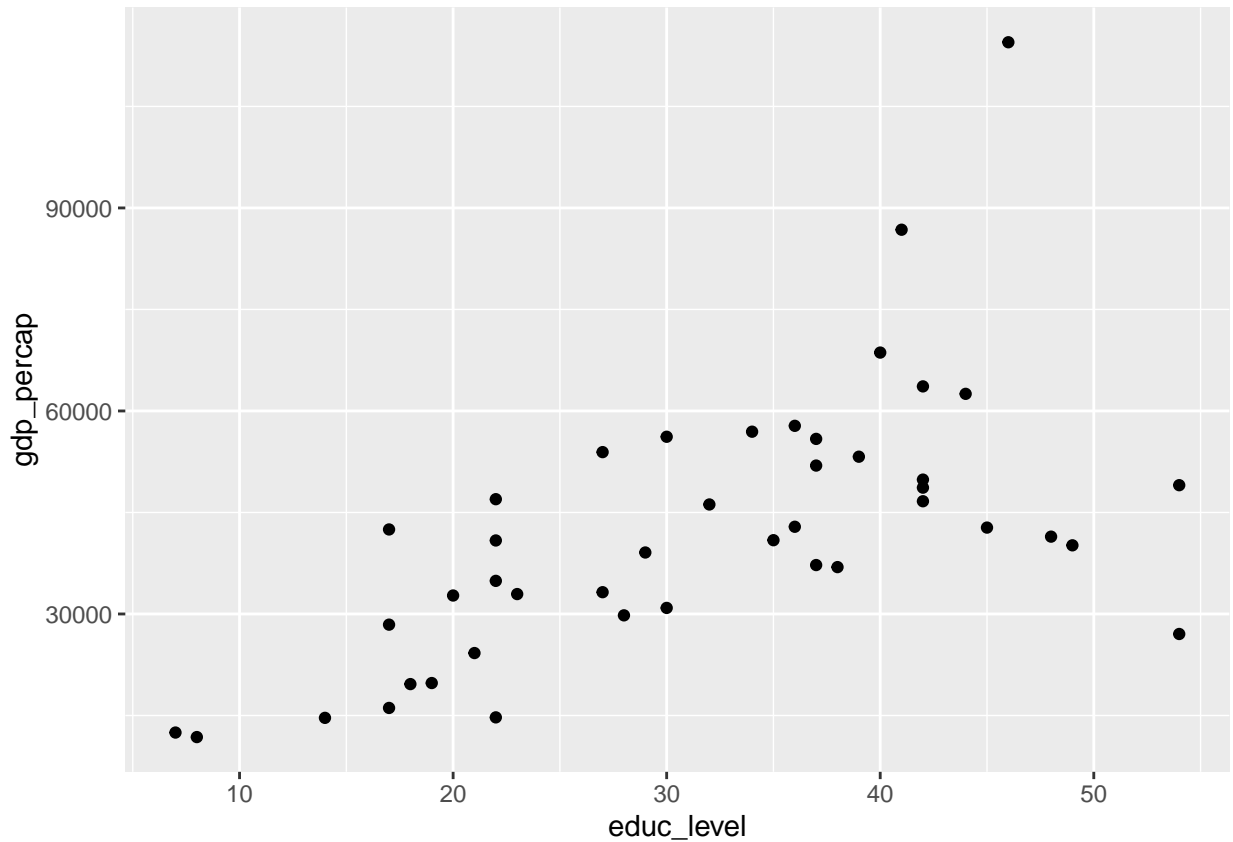
```
gdp_percap$country <- gsub('.{2}$', '', gdp_percap$country)

combined<-left_join(educ,gdp_percap,by="country")
```

4. Plot gdp per capita (CIA version) as a function of the percent of the population aged 25-64 with a tertiary education.

```
gg<-combined%>%
  ggplot(aes(x=educ_level,y=gdp_percap,label=country))+
  geom_point()

#ggplotly(gg)
gg
```



5. Model gdp per capita (using a linear model) as a function of the percent of the population aged 25-64 with a tertiary education.

```
mod1<-lm(gdp_percap~educ_level,data=combined)
summary(mod1)

##
## Call:
## lm(formula = gdp_percap ~ educ_level, data = combined)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -37698  -9728  -3292    8239   57782
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10456.2    6757.1    1.547   0.129
```

```
## educ_level    1005.3      200.3   5.019    1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15720 on 42 degrees of freedom
## Multiple R-squared:  0.3749, Adjusted R-squared:  0.36
## F-statistic: 25.19 on 1 and 42 DF,  p-value: 1.001e-05
```