DACSS 601 Final Project

Global Disparities in Internet Access

Alphoncina Lyamuya

8/7/2021

Contents

Introduction

Internet access is one of the 21st-century essentials to almost all spheres of life. Despite the push to make the internet a public utility in most advanced economies as well as the 2016 UN General Assembly passage of a non-binding Resolution declaring "Internet access a human right," Internet access has remained uneven across the globe (Assembly 2016; Brendan 2014). The International Telecommunication Union (ITU) reported that by the end of 2016, there were 3.9 billion people that were not using the Internet (International Telecommunication Union 2016). With a total population of 7.4 billion in 2016, this implies about 53% of the world's population was digitally excluded (World Bank, n.d.; International Telecommunication Union 2016). When a significant portion of the population of the Global North countries has been reported as active Internet users, many in the Global South are digitally excluded. For instance, when only 21% of the European population was reported as non-users in 2016, about 75% of the population in Africa was still not using the Internet (International Telecommunication Union 2016).

Most people access the internet through a certain kind of broadband connection. Also referred to as high-speed internet access, broadband includes the following types of connections: DSL (Digital Subscriber Line), fiber-optic, cable, satellite, and wireless (Federal Communications Commission 2014; International Telecommunication Union 2003)). When DSL, fiber-optic, cable and satellite connections are categorized as fixed broadband, "wireless broadband can be mobile or fixed" (Federal Communications Commission 2014). By 2017, there were about 7.75 billion mobile broadband subscriptions and about 1.02 billion fixed broadband subscriptions (World Bank, n.d.). With the Covid-19 pandemic and the ensued global shutdowns, access to reliable high-speed Internet has become more than ever important. From education, to health services, to jobs and the economy at large, fixed broadband internet has emerged as the cornerstone to social interactions, economic functionality, and even public health responses. But this has largely been relevant to those that can afford the cost of high-speed internet and live in regions or places with functioning Internet infrastructure. Given the growing need for high-speed Internet, the existing disparities in Internet access, and the socio-economic potential of fixed broadband connection, this study asks: to what extent does a country's per capita GDP corresponds to access to fixed broadband internet? I hypothesized that countries with higher per capita GDP are more likely to have well-established internet physical infrastructure as well as high access to percapelectity and high median income which allow most of their population to have access to fixed broadband Internet. The findings from this study would be useful in informing international policymakers and stakeholders in their effort to design effective measures to address different layers of the global digital divide.

Data

The dataset used for this project is a compilation of several secondary datasets all from the World Bank Open Data. These datasets include data on fixed broadband subscriptions from 177 countries for the year 2007 to 2017; data on per capita GDP from 177 countries for the year 2007 to 2017; data on percapelectity access from 177 countries for the year 2007 to 2017; data on population from 177 countries for the year 2007 to 2017; and data on cellular mobile subscriptions from 177 countries for the year 2007 to 2017. The compiled dataset included the following variable: *country* (name of the country), *year* (period covered in the data set, 2007-2017), *population* (total population of a country in a given year), *fixbroads* (fixed broadband subscriptions in a given year), *percapelect* (percapelectity access across a country in a given year), *gdpercap* (per capita GDP of a country in a given year), *mobsub* (cellular mobile subscriptions in a country). This dataset was then analyzed using R statistical software to examine the statistical relationship between the dependent variable (fixed broadband subscriptions) and independent variable (per capita GDP).

Import Data and Run Libraries

```
library(readr)
library(tidyverse)
library(tidyr)
library(ggplot2)
library(psych)
```

```
# Import the dataset
myData <- read.csv("newData.csv")
myData1 <- myData

# Snapshot of the dataset
head(myData)
```

```
##   country_name year gdpercap fixbroads mobsub  electric population
## 1        Aruba 2007 25834.11     15996 113586 100.00000     101226
## 2        Aruba 2008 27083.63     18396 120806 100.00000     101362
## 3        Aruba 2009 24631.18     18800 128000 100.00000     101452
## 4        Aruba 2010 23513.53     19217 131800  93.35629     101665
## 5        Aruba 2011 24985.01        NA     NA 100.00000     102050
## 6        Aruba 2012 24712.49        NA 135000 100.00000     102565
```

```
# Description of the data
glimpse(myData)
```

```
## Rows: 1,947
## Columns: 7
## $ country_name <fct> "Aruba", "Aruba", "Aruba", "Aruba", "Aruba", "Aruba", "Ar~
## $ year         <int> 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 201~
## $ gdpercap     <dbl> 25834.1118, 27083.6349, 24631.1821, 23513.5277, 24985.013~
## $ fixbroads    <int> 15996, 18396, 18800, 19217, NA, NA, 19200, 19200, 19000, ~
## $ mobsub       <dbl> 113586, 120806, 128000, 131800, NA, 135000, 138800, 13970~
## $ electric     <dbl> 100.00000, 100.00000, 100.00000, 93.35629, 100.00000, 100~
## $ population   <int> 101226, 101362, 101452, 101665, 102050, 102565, 103165, 1~
```

Variable Description

The dependent variable (DV) fixed broadband subscription (`fixbroads`) represent total residential and organizational subscriptions to high-speed access to the Internet "at downstream speeds equal to, or greater than, 256 kbit/s. This includes cable modem, DSL, fiber-to-the-home/building, other fixed (wired)-broadband subscriptions, satellite broadband" (`World bank`).

The independent variable (IV) per capita GDP (`gdpercap`) represents the gross domestic product per midyear population of a country in a given year.

```
# Describing the variable
str(myData$fixbroads)
```

```
##  int [1:1947] 15996 18396 18800 19217 NA NA 19200 19200 19000 NA ...
```

```
str(myData$gdpercap)
```

```
##  num [1:1947] 25834 27084 24631 23514 24985 ...
```

```
summary(myData$fixbroads)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.      NA's
##         0     18295    197000   3820911   1770574 394190000        78
```

```
summary(myData$gdpercap)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
##    172.5   1660.6   5576.8  16205.7  20426.3 189432.4        8
```

Data Cleaning and Transformation

```
# Check for missing values
table(is.na(myData))
```

```
##
## FALSE  TRUE
## 13526   103
```

```r
# Examine the variables with missing values
myData[rowSums(is.na(myData)) > 0, ]

table(is.na(myData$aliteracy))
```

- There are **103** missing values in the dataset `myData`. These missing values are found across `fixbroads`, `gdpercap`, `mobsub`, and `percapelect`. To avoid drawing inaccurate inference about the data, and given their relatively smaller size compared to the total values in the dataset, I will not remove existing missing values from the dataset.

Data Transformation

Here some variable will be renamed, some new ones will be created, and the dataset will be arranged by names of countries (`country`).

```r
# Transforming variables
myDataNew <- myData %>%
  rename("country" = "country_name") %>%
  mutate(percapfbroad = fixbroads/population,
         percapmobsub = mobsub/population,
         percapelect = electric/population) %>%
  arrange(country)

head(myDataNew)
```

```
##       country year gdpercap fixbroads   mobsub electric population percapfbroad
## 1 Afghanistan 2007 359.6932       500  4668096 33.90160   27100542 1.844982e-05
## 2 Afghanistan 2008 364.6607       500  7898909 42.40000   27722281 1.803603e-05
## 3 Afghanistan 2009 438.0761      1000 10500000 45.52068   28394806 3.521771e-05
## 4 Afghanistan 2010 543.3030      1500 10215840 42.70000   29185511 5.139537e-05
## 5 Afghanistan 2011 591.1628        NA 13797879 43.22202   30117411           NA
## 6 Afghanistan 2012 641.8714      1500 15340115 69.10000   31161378 4.813651e-05
##   percapmobsub  percapelect
## 1    0.1722510 1.250957e-06
## 2    0.2849300 1.529456e-06
## 3    0.3697859 1.603134e-06
## 4    0.3500312 1.463055e-06
## 5    0.4581363 1.435117e-06
## 6    0.4922797 2.217489e-06
```

```r
# Filtering data by year
for (year in myDataNew){
  filter2007 <- filter(myDataNew, year == 2007)
  filter2008 <- filter(myDataNew, year == 2008)
  filter2009 <- filter(myDataNew, year == 2009)
  filter2010 <- filter(myDataNew, year == 2010)
  filter2011 <- filter(myDataNew, year == 2011)
  filter2012 <- filter(myDataNew, year == 2012)
  filter2013 <- filter(myDataNew, year == 2013)
  filter2014 <- filter(myDataNew, year == 2014)
  filter2015 <- filter(myDataNew, year == 2015)
  filter2016 <- filter(myDataNew, year == 2016)
```

```
  filter2017 <- filter(myDataNew, year == 2017)
}
```

Summary Description and Data Exploration

After creating a variable `percapfbroad` (fixed broadband subscription (per 100 people)) from the variable `fixbroads`, `percapfbroad` will be used as a DV throughout the rest of the analysis.

```
options(scipen = 999, digits = 3)

# Exploration of measures of central tendencies and distribution of the variable
describe(myDataNew$percapfbroad)
```

```
##    vars    n mean   sd median trimmed  mad min  max range skew kurtosis se
## X1    1 1869 0.11 0.13   0.05    0.09 0.07   0 0.62  0.62 1.02    -0.01  0
```

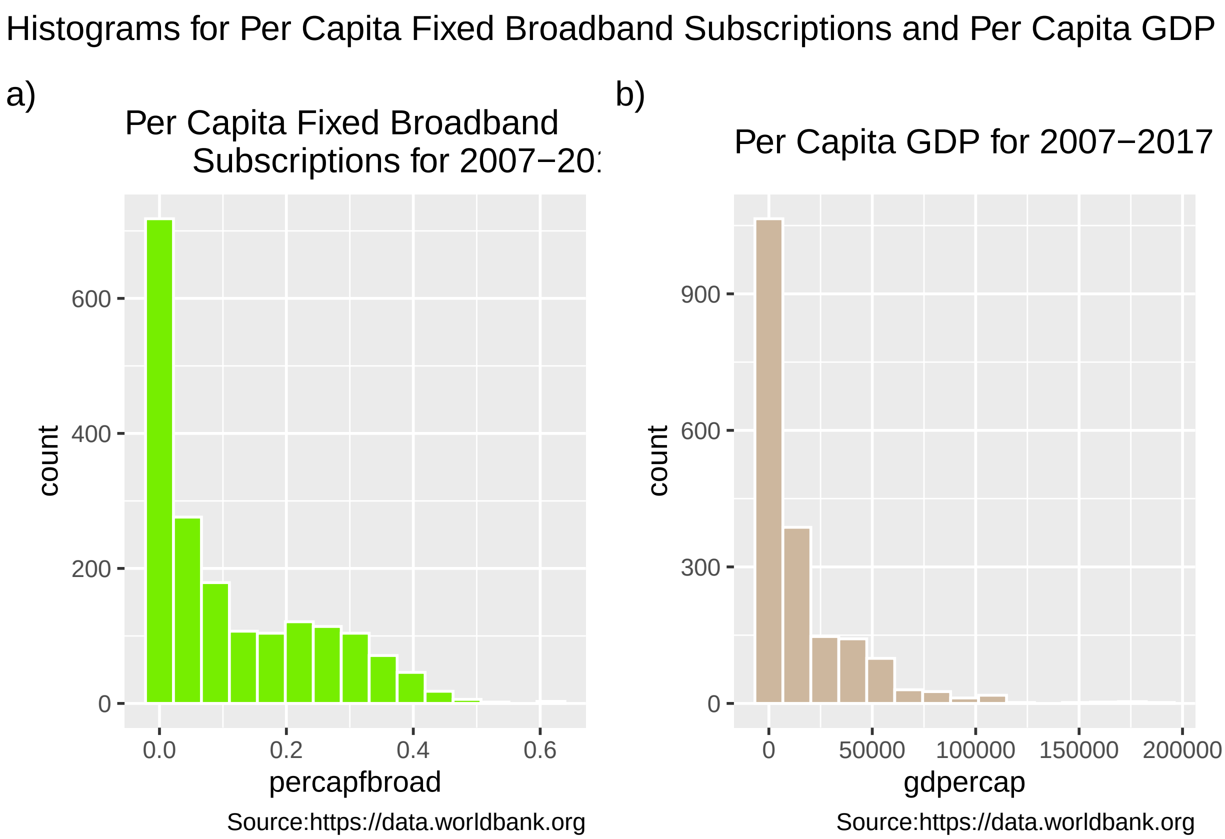
```
describe(myDataNew$gdpercap)
```

```
##    vars    n  mean    sd median trimmed  mad min    max  range skew kurtosis
## X1    1 1939 16206 24234   5577   11041 6890 172 189432 189260 2.77     10.4
##     se
## X1 550
```

```
# Visual distribution of the variables
plot1a <- ggplot(data = myDataNew, mapping = aes(x = percapfbroad)) +
  geom_histogram(fill = "chartreuse2", bins = 15, color = "white") +
  labs(title = "Per Capita Fixed Broadband
       Subscriptions for 2007-2017",
     caption = "Source:https://data.worldbank.org")

plot1b <- ggplot(data = myDataNew, mapping = aes(x = gdpercap)) +
  geom_histogram(fill = "bisque3", bins = 15, color = "white") +
  labs(title = "Per Capita GDP for 2007-2017",
     caption = "Source:https://data.worldbank.org")

# arrange the two plots above
library(patchwork)

plot1a + plot1b + plot_annotation(title =
"Histograms for Per Capita Fixed Broadband Subscriptions and Per Capita GDP",
                                tag_levels="a", tag_suffix = ")")
```

```
## Warning: Removed 78 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 8 rows containing non-finite values (stat_bin).
```

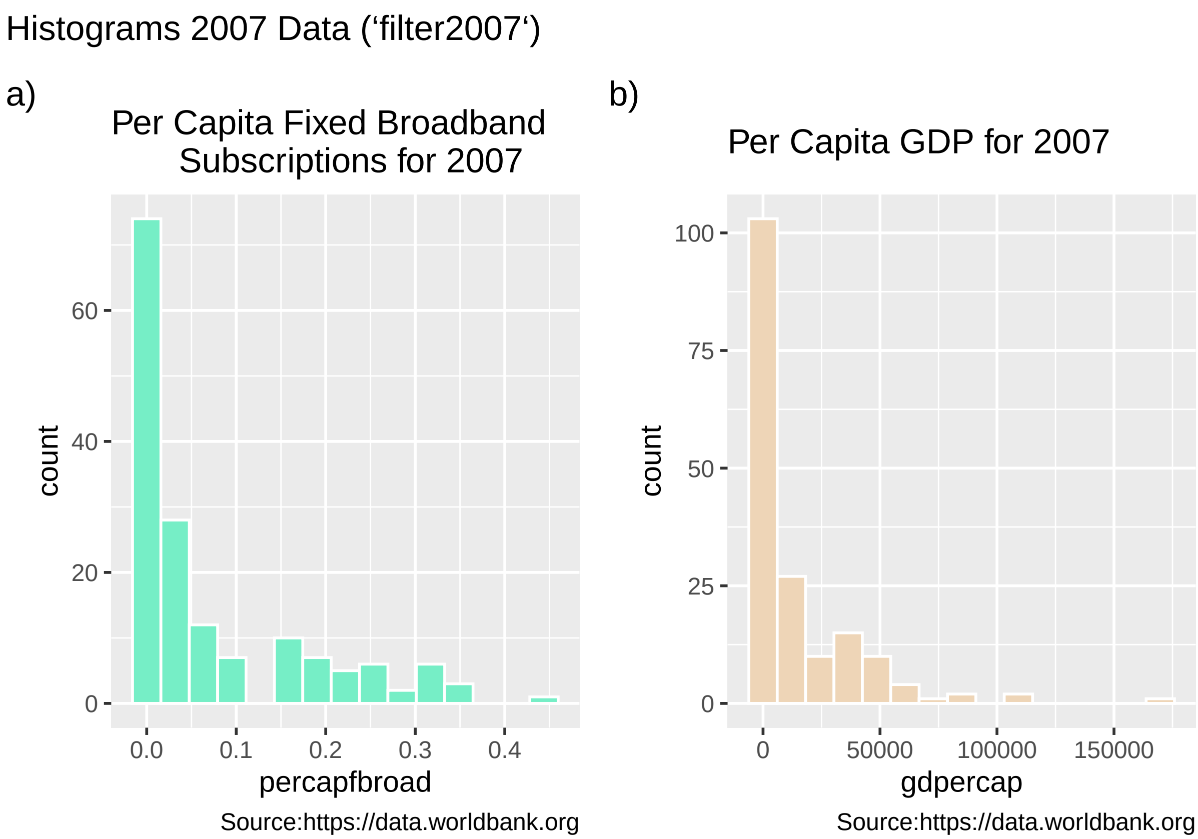


To further explore the distribution of the two variables, `gdpercap` and `percapfbroad`, the filtered data based on individual year of observation will be explored. Given there are 11 sets of filtered data, the `filter2007`, `filtered2012`, and `filtered2017` datasets will be used as example to demonstrate the distribution of the DV and IV.

```
# 2007 data
plot2a <- ggplot(data = filter2007, mapping = aes(x = percapfbroad)) +
  geom_histogram(fill = "aquamarine2", bins = 15, color = "white") +
  labs(title = "Per Capita Fixed Broadband
       Subscriptions for 2007",
     caption = "Source:https://data.worldbank.org")

plot2b <- ggplot(data = filter2007, mapping = aes(x = gdpercap)) +
  geom_histogram(fill = "bisque2", bins = 15, color = "white") +
  labs(title = "Per Capita GDP for 2007",
     caption = "Source:https://data.worldbank.org")

plot2a + plot2b + plot_annotation(title = "Histograms 2007 Data (`filter2007`)",
                                  tag_levels="a", tag_suffix = ")")
```

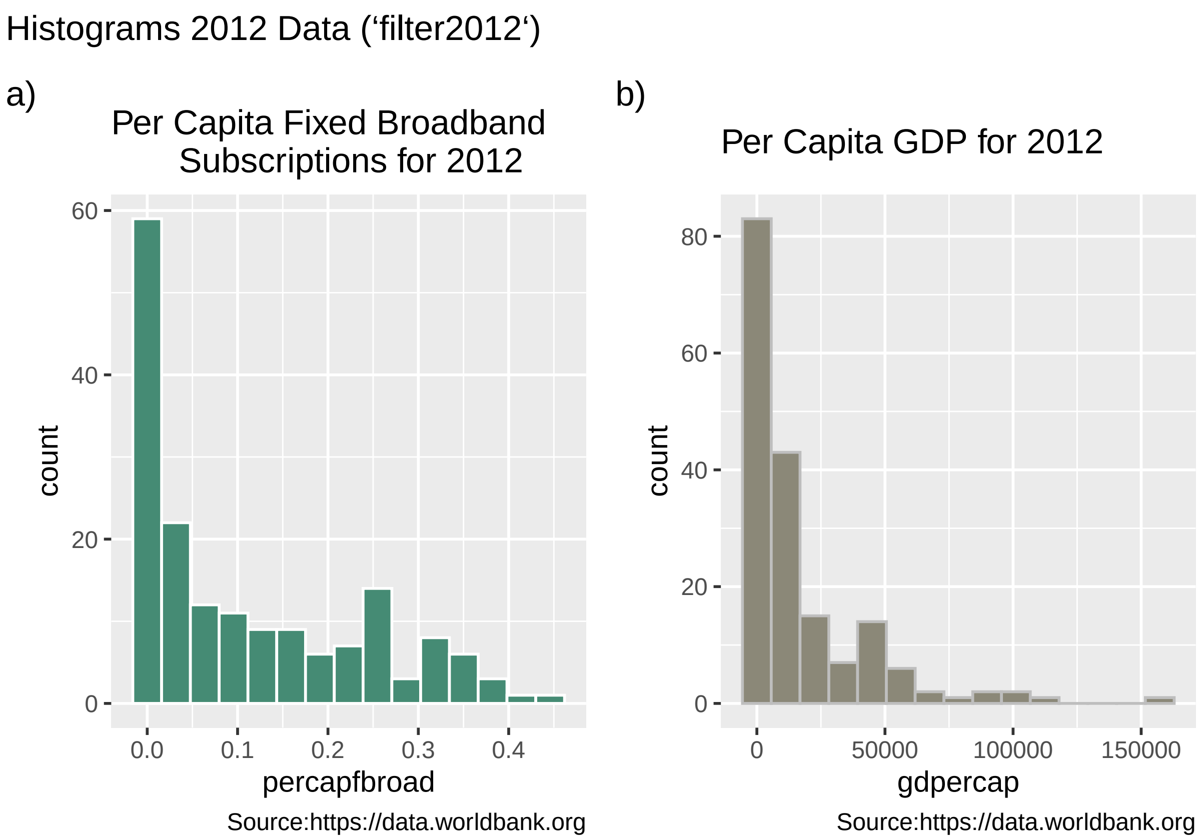
```
## Warning: Removed 16 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 2 rows containing non-finite values (stat_bin).
```




```r
# 2012 data
plot3a <- ggplot(data = filter2012, mapping = aes(x = percapfbroad)) +
  geom_histogram(fill = "aquamarine4", bins = 15, color = "white") +
  labs(title = "Per Capita Fixed Broadband
       Subscriptions for 2012",
     caption = "Source:https://data.worldbank.org")

plot3b <- ggplot(data = filter2012, mapping = aes(x = gdpercap)) +
  geom_histogram(fill = "cornsilk4", bins = 15, color = "grey") +
  labs(title = "Per Capita GDP for 2012",
     caption = "Source:https://data.worldbank.org")

plot3a + plot3b + plot_annotation(title = "Histograms 2012 Data (`filter2012`)",
                                tag_levels="a", tag_suffix = ")")
```

```
## Warning: Removed 6 rows containing non-finite values (stat_bin).
```

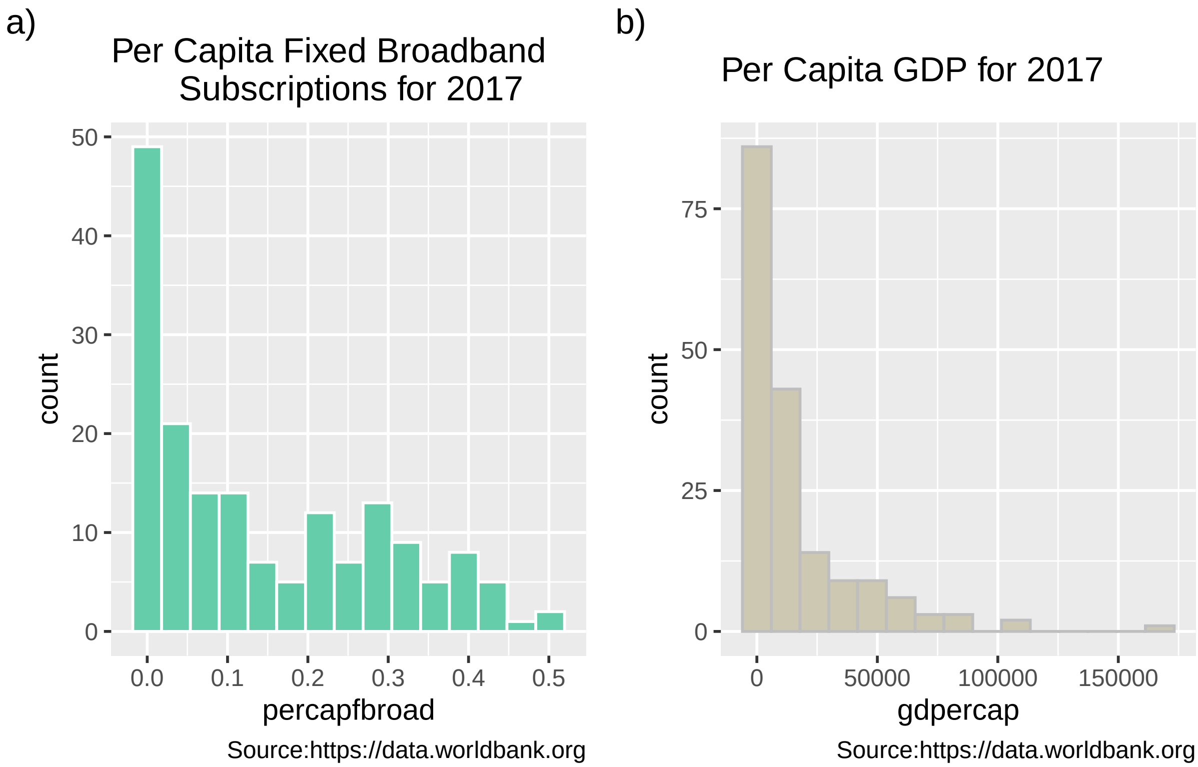



```
# 2017 data
plot4a <- ggplot(data = filter2017, mapping = aes(x = percapfbroad)) +
  geom_histogram(fill = "aquamarine3", bins = 15, color = "white") +
  labs(title = "Per Capita Fixed Broadband
       Subscriptions for 2017",
     caption = "Source:https://data.worldbank.org")

plot4b <- ggplot(data = filter2017, mapping = aes(x = gdpercap)) +
  geom_histogram(fill = "cornsilk3", bins = 15, color = "grey") +
  labs(title = "Per Capita GDP for 2017",
     caption = "Source:https://data.worldbank.org")

plot4a + plot4b + plot_annotation(title = "Histograms 2017 Data (`filter2017`)",
                                tag_levels="a", tag_suffix = ")")
```

```
## Warning: Removed 5 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

Histograms 2017 Data (‘filter2017‘)



Visualization

Correlation and Regression Analysis

In this anaysis, I'm examining the statistical relationship between `gdpercap` (IV) and `percapfbroad` (DV). To further analyze how the two variables relate to one another, the filtered dataset by year would be examined to explore how over time the relationship between the two variable change. This will be useful at not only answering the research question but also to demonstrate the effect of time of the relationship between the IV and DV.

Furthermore, other variables including `percapelect` and `percapmobsub` will be included in the some of the statistical models to examine whether these factors have a moderator effect on the relationship between the IV and the DV.

This is the description of my null and alternative hypothesis:

- H_0 (*null hypothesis*): High per capita GDP does not correspond to a high per capita fixed broadband subscription in a country.
- H_1 (*alternative hypothesis*): High per capita GDP does correspond to a high per capita fixed broadband subscription in a country.

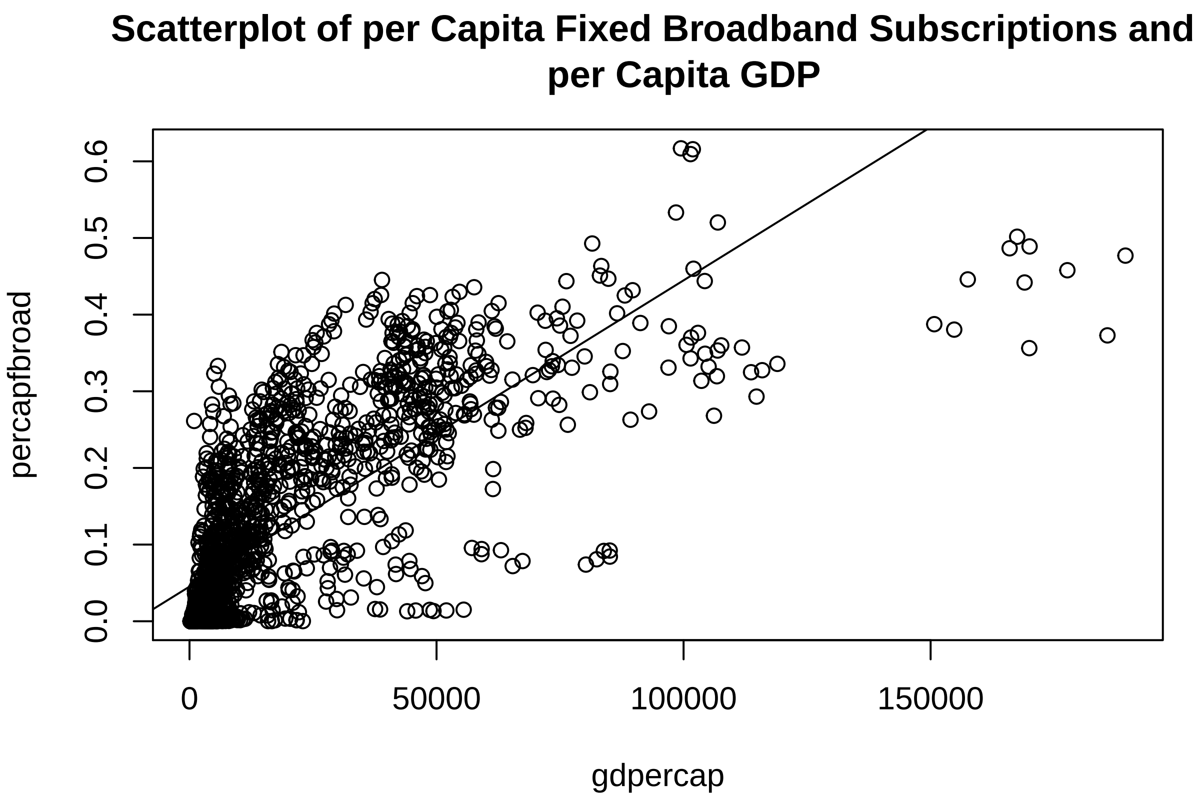
```r
options(scipen = 999, digits = 3)

# Correlation
cor1 <- cor(myDataNew$gdpercap, myDataNew$percapfbroad,
            use="pairwise.complete.obs")
cor1
```

```
## [1] 0.766
```

```r
# Bivariate Regression
lm1 <- lm(percapfbroad ~ gdpercap, data = myDataNew)
summary(lm1)
```

```
##
## Call:
## lm(formula = percapfbroad ~ gdpercap, data = myDataNew)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4145 -0.0476 -0.0285  0.0482  0.2648
##
## Coefficients:
##                 Estimate   Std. Error t value            Pr(>|t|)
## (Intercept) 0.0451770144 0.0022923663    19.7 <0.0000000000000002 ***
## gdpercap    0.0000039954 0.0000000778    51.3 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0818 on 1859 degrees of freedom
##   (86 observations deleted due to missingness)
## Multiple R-squared:  0.586,  Adjusted R-squared:  0.586
## F-statistic: 2.63e+03 on 1 and 1859 DF,  p-value: <0.0000000000000002
```

```r
# Plot the relationship
plot(percapfbroad ~ gdpercap,
       data = myDataNew,
       main = "Scatterplot of per Capita Fixed Broadband Subscriptions and
     per Capita GDP")

# add the regression line
abline(lm1)
```



For cor1

- There is a positive correlation of **0.766**. That is there is a positive relationship between per capita GDP and per capita fixed broadband subscription across countries.

For model lm(1)

- The p-value is less than 0.05. That is, the relationship between country's per capita GDP and per capita fixed broadband subscription is significant at 0.05 level. Therefore, there is evidence to reject the null hypothesis for the alternative hypothesis at 0.05 level.

```
# Test the correlation or the filtered dataset for 2007, 2009, 2011, 2013, 2015,
# and 2017

(cor2 <- cor(filter2007$gdpercap, filter2007$percapfbroad,
            use="pairwise.complete.obs"))
```

```
## [1] 0.833
```

```
(cor3 <- cor(filter2009$gdpercap, filter2009$percapfbroad,
            use="pairwise.complete.obs"))
```

```
## [1] 0.837
```

```
(cor4 <- cor(filter2011$gdpercap, filter2011$percapfbroad,
           use="pairwise.complete.obs"))
```

```
## [1] 0.777
```

```
(cor5 <- cor(filter2013$gdpercap, filter2013$percapfbroad,
           use="pairwise.complete.obs"))
```

```
## [1] 0.761
```

```
(cor6 <- cor(filter2015$gdpercap, filter2015$percapfbroad,
           use="pairwise.complete.obs"))
```

```
## [1] 0.765
```

```
(cor7 <- cor(filter2017$gdpercap, filter2017$percapfbroad,
           use="pairwise.complete.obs"))
```

```
## [1] 0.752
```

For cor2

- There is a positive correlation of **0.833**. That is there is a positive relationship between per capita GDP and per capita fixed broadband subscription across countries.

For cor3

- There is a positive correlation of **0.837**. That is there is a positive relationship between per capita GDP and per capita fixed broadband subscription across countries.

For cor4

- There is a positive correlation of **0.777**. That is there is a positive relationship between per capita GDP and per capita fixed broadband subscription across countries.

For cor5

- There is a positive correlation of **0.761**. That is there is a positive relationship between per capita GDP and per capita fixed broadband subscription across countries.

For cor6

- There is a positive correlation of **0.765**. That is there is a positive relationship between per capita GDP and per capita fixed broadband subscription across countries.

For cor7

- There is a positive correlation of **0.752**. That is there is a positive relationship between per capita GDP and per capita fixed broadband subscription across countries.

```r
# Regression for the filtered dataset for 2007, 2009, 2011, 2013, 2015, and 2017

lm2 <- lm(percapfbroad ~ gdpercap, data = filter2007)
summary(lm2)
```

```
##
## Call:
## lm(formula = percapfbroad ~ gdpercap, data = filter2007)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2636 -0.0200 -0.0168  0.0241  0.2001
##
## Coefficients:
##                Estimate  Std. Error t value            Pr(>|t|)
## (Intercept) 0.016514249 0.005407677    3.05              0.0027 **
## gdpercap    0.000003550 0.000000188   18.89 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0569 on 157 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.694,  Adjusted R-squared:  0.692
## F-statistic:  357 on 1 and 157 DF,  p-value: <0.0000000000000002
```

```r
lm3 <- lm(percapfbroad ~ gdpercap, data = filter2009)
summary(lm3)
```

```
##
## Call:
## lm(formula = percapfbroad ~ gdpercap, data = filter2009)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3175 -0.0297 -0.0199  0.0307  0.2218
##
## Coefficients:
##                Estimate  Std. Error t value             Pr(>|t|)
## (Intercept) 0.026691054 0.005992108    4.45             0.000015 ***
## gdpercap    0.000004337 0.000000221   19.67 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0643 on 166 degrees of freedom
##   (9 observations deleted due to missingness)
## Multiple R-squared:   0.7,  Adjusted R-squared:  0.698
## F-statistic:  387 on 1 and 166 DF,  p-value: <0.0000000000000002
```

```r
lm4 <- lm(percapfbroad ~ gdpercap, data = filter2011)
summary(lm4)
```

```
##
```

```
## Call:
## lm(formula = percapfbroad ~ gdpercap, data = filter2011)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2689 -0.0425 -0.0211  0.0464  0.2230
##
## Coefficients:
##                Estimate  Std. Error t value             Pr(>|t|)
## (Intercept) 0.040525558 0.007004075    5.79          0.000000035 ***
## gdpercap    0.000003752 0.000000235   15.96 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.075 on 167 degrees of freedom
##   (8 observations deleted due to missingness)
## Multiple R-squared:  0.604,  Adjusted R-squared:  0.602
## F-statistic:  255 on 1 and 167 DF,  p-value: <0.0000000000000002
```

```r
lm5 <- lm(percapfbroad ~ gdpercap, data = filter2013)
summary(lm5)
```

```
##
## Call:
## lm(formula = percapfbroad ~ gdpercap, data = filter2013)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2886 -0.0541 -0.0245  0.0689  0.2143
##
## Coefficients:
##                Estimate  Std. Error t value             Pr(>|t|)
## (Intercept) 0.052220938 0.007882796    6.62        0.00000000044 ***
## gdpercap    0.000003863 0.000000253   15.29 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0852 on 170 degrees of freedom
##   (5 observations deleted due to missingness)
## Multiple R-squared:  0.579,  Adjusted R-squared:  0.577
## F-statistic:  234 on 1 and 170 DF,  p-value: <0.0000000000000002
```

```r
lm6 <- lm(percapfbroad ~ gdpercap, data = filter2015)
summary(lm6)
```

```
##
## Call:
## lm(formula = percapfbroad ~ gdpercap, data = filter2015)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3110 -0.0621 -0.0282  0.0657  0.2190
##
```

```
## Coefficients:
##                Estimate  Std. Error t value             Pr(>|t|)
## (Intercept) 0.060554017 0.008133109    7.45      0.0000000000046 ***
## gdpercap    0.000004440 0.000000287   15.49 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.088 on 170 degrees of freedom
##   (5 observations deleted due to missingness)
## Multiple R-squared:  0.585,  Adjusted R-squared:  0.583
## F-statistic:  240 on 1 and 170 DF,  p-value: <0.0000000000000002
```

```r
lm7 <- lm(percapfbroad ~ gdpercap, data = filter2017)
summary(lm7)
```

```
##
## Call:
## lm(formula = percapfbroad ~ gdpercap, data = filter2017)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2965 -0.0721 -0.0285  0.0705  0.2376
##
## Coefficients:
##                Estimate  Std. Error t value             Pr(>|t|)
## (Intercept) 0.070382246 0.008779843    8.02     0.00000000000017 ***
## gdpercap    0.000004344 0.000000293   14.82 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0946 on 169 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.565,  Adjusted R-squared:  0.563
## F-statistic:  220 on 1 and 169 DF,  p-value: <0.0000000000000002
```

```r
# Test whether there is correlation between DV, IV, and other IVS
lm8 <- lm(percapfbroad ~ gdpercap + percapelect, data = myDataNew)
summary(lm8)
```

```
##
## Call:
## lm(formula = percapfbroad ~ gdpercap + percapelect, data = myDataNew)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4030 -0.0484 -0.0282  0.0478  0.2639
##
## Coefficients:
##                  Estimate    Std. Error t value            Pr(>|t|)
## (Intercept)  0.0459864622  0.0023014115   19.98 <0.0000000000000002 ***
## gdpercap     0.0000040337  0.0000000791   50.99 <0.0000000000000002 ***
## percapelect -6.7797122685  2.4961879148   -2.72              0.0067 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0817 on 1847 degrees of freedom
##   (97 observations deleted due to missingness)
## Multiple R-squared:  0.589,  Adjusted R-squared:  0.588
## F-statistic: 1.32e+03 on 2 and 1847 DF,  p-value: <0.0000000000000002
```

```
lm9 <- lm(percapfbroad ~ gdpercap + percapelect + percapmobsub, data = myDataNew)
summary(lm9)
```

```
##
## Call:
## lm(formula = percapfbroad ~ gdpercap + percapelect + percapmobsub,
##     data = myDataNew)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3129 -0.0411 -0.0147  0.0382  0.2516
##
## Coefficients:
##                    Estimate    Std. Error t value             Pr(>|t|)
## (Intercept)  -0.0150961742  0.0045494964   -3.32              0.00092 ***
## gdpercap      0.0000035643  0.0000000809   44.07 < 0.0000000000000002 ***
## percapelect  -1.7080808355  2.4789878457   -0.69              0.49090
## percapmobsub  0.0691037875  0.0045244990   15.27 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.077 on 1845 degrees of freedom
##   (98 observations deleted due to missingness)
## Multiple R-squared:  0.635,  Adjusted R-squared:  0.634
## F-statistic: 1.07e+03 on 3 and 1845 DF,  p-value: <0.0000000000000002
```

For model lm(2) to model lm(7)

- The p-value is less than 0.05. That is, the relationship between country's per capita GDP and per capita fixed broadband subscription is significant at 0.05 level in the year 2007, 2009, 2011, 2013, 2015, 2017. Therefore, there is evidence to reject the null hypothesis for the alternative hypothesis at 0.05 level.

For model lm(8)

- `gdpercap`: The p-value is significant at 0.05 level.
- `percapelect`: The p-value of (0.0067) is not statistically significant at 0.05 level

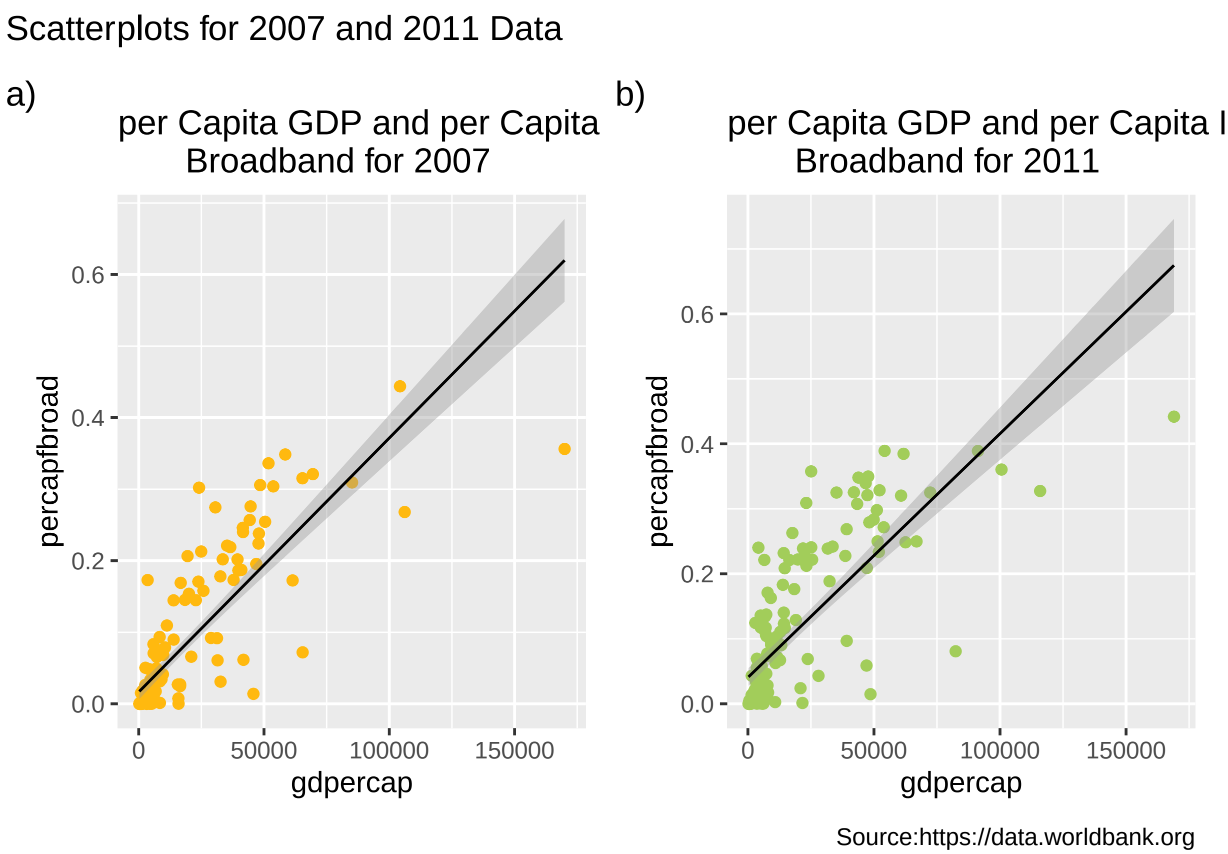
For model lm(9)

- `gdpercap`: The p-value is significant at 0.05 level.
- `percapmobsub`: The p-value is significant at 0.05 level.
- `percapelect`: The p-value of (0.49090) is not statistically significant at 0.05 level

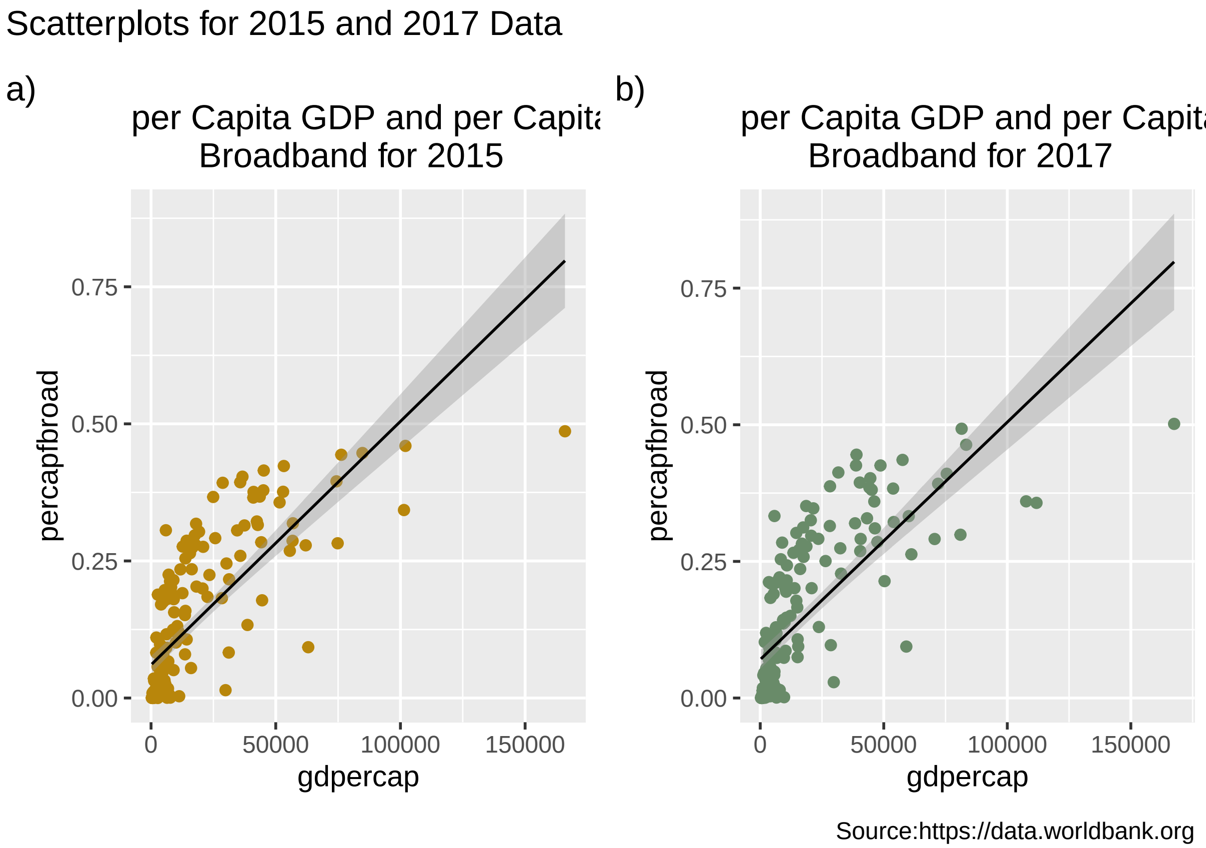
---

GG-Plot Data Visualization

```
# Scatterplot for `percapfbroad` and `gdpercap`
plot5 <- filter2007 %>%
  ggplot(mapping = aes(x = gdpercap, y = percapfbroad)) +
  geom_point(color = "darkgoldenrod1") +
  geom_smooth(method = "lm", color = "black", size = 0.5) +
  labs(title = "per Capita GDP and per Capita Fixed
       Broadband for 2007")


plot6 <- filter2011 %>%
  ggplot(mapping = aes(x = gdpercap, y = percapfbroad)) +
  geom_point(color = "darkolivegreen3") +
  geom_smooth(method = "lm", color = "black", size = 0.5) +
  labs(title = "per Capita GDP and per Capita Fixed
       Broadband for 2011")


plot7 <- filter2015 %>%
  ggplot(mapping = aes(x = gdpercap, y = percapfbroad)) +
  geom_point(color = "darkgoldenrod") +
  geom_smooth(method = "lm", color = "black", size = 0.5) +
  labs(title = "per Capita GDP and per Capita Fixed
       Broadband for 2015")

plot8 <- filter2017 %>%
  ggplot(mapping = aes(x = gdpercap, y = percapfbroad)) +
  geom_point(color = "darkseagreen4") +
  geom_smooth(method = "lm", color = "black", size = 0.5) +
  labs(title = "per Capita GDP and per Capita Fixed
       Broadband for 2017")

# Arrange the plots
plot5 + plot6 + plot_annotation(title = "Scatterplots for 2007 and 2011 Data",
                                tag_levels="a", tag_suffix = ")",
                                caption = "Source:https://data.worldbank.org")
```




```r
plot7 + plot8 + plot_annotation(title = "Scatterplots for 2015 and 2017 Data",
                                tag_levels="a", tag_suffix = ")",
                                caption = "Source:https://data.worldbank.org")
```



Reflection

Going through the process of data collection and compilation not only helped me to learn and improve my analytical skills but also taught me to appreciate the work that goes behind any statistical analysis. Drawing from the dataset and some of the variables, there were several constraints. First, I had to reconstruct the dataset by compiling several datasets from the World Bank to have all my variables of interest in a single dataset. Second, the variable fixed broadband subscriptions included both residential and organizational subscriptions. This poses some challenges in analyzing the distribution of those subscriptions to understand how many households in a country had access to high-speed internet at their places of residence.

Moving on to data wrangling and exploration, I encountered several challenges in variable transformation and in making analytical choices around the handling of observations across all 11 years. Given that I was using a panel dataset with data from 2007 to 2017, I wish I was able to perform some more advanced regression analysis to examine the effect of time on the relationship between my IV and DV.

If I were to continue the project, I would examine the effect of fixed broadband internet on countries' economic growth and development. The dataset used in this project could also be useful in examining such an effect.

Conclusion

In brief, this paper examined whether countries' per capita GDP influences people's access to fixed broadband internet. Data from 2007 to 2017 was used in the examination of this research question. To analyze whether per capita GDP (IV) had an influence on per capita fixed broadband subscriptions (DV) in a country, I run a correlation between the two variables based on the entire dataset with data from 2007 to 2017. I found a positive correlation of 0.776 between the IV and DV. I went further to examine the relationship between the variable based on data from an individual year. Six correlations were run, and they all revealed a positive coefficient, indicating a positive relationship between per capita GDP and per capita fixed broadband subscription across the six years included. There was an incremental change in the correlation coefficient from 2007 to 2011, however, from 2011 to 2017 there was a decremental change in the correlation coefficient. Given that correlation does not mean causality, I went on to conduct linear regression analysis.

For regression analysis, I run 9 linear regression models. While models (lm1), lm(8), and lm(9) used the large dataset with all 11 years of observations, models lm(2) to lm(7) were based on the filtered dataset with 177 observations in a particular year (2007, 2009, 2011, 2013, 2015, and 2017). For models lm(1) to lm(7), the p-values were less than 0.05. That is, the relationship between a country's per capita GDP and per capita fixed broadband subscription was statistically significant at 0.05 level in the years 2007, 2009, 2011, 2013, 2015, 2017 as well as the overall period between 2007 to 2017. Thus, there was evidence to reject the null hypothesis for the alternative hypothesis.

In model lm(8), I regressed per capita fixed broadband subscription on per capita GDP and per capita electricity. While the p-values for per capita GDP was statistical significant at 0.05 level, that of per capita electricity was not statistically significant. In model lm(9), I regressed per capita fixed broadband subscription on per capita GDP, per capita electricity, and per capita cellular mobile subscription. When the p-values for both per capita GDP and per capita cellular mobile subscription were statistically significant at 0.05 level, the p-value for per capita electricity was not statistically significant at 0.05 level.

The results from all regression models are summarized in the table below.

Model Summary

```
library(sjmisc)
```

```
##
## Attaching package: 'sjmisc'

## The following object is masked from 'package:purrr':
##
##     is_empty

## The following object is masked from 'package:tidyr':
##
##     replace_na

## The following object is masked from 'package:tibble':
##
##     add_case
```

```
library(sjPlot)
```

```
## Registered S3 methods overwritten by 'parameters':
##   method                           from
##   as.double.parameters_kurtosis    datawizard
##   as.double.parameters_skewness    datawizard
##   as.double.parameters_smoothness  datawizard
```

```
##   as.numeric.parameters_kurtosis   datawizard
##   as.numeric.parameters_skewness   datawizard
##   as.numeric.parameters_smoothness datawizard
##   print.parameters_distribution    datawizard
##   print.parameters_kurtosis        datawizard
##   print.parameters_skewness        datawizard
##   summary.parameters_kurtosis      datawizard
##   summary.parameters_skewness      datawizard
```

```
library(sjlabelled)
```

```
##
## Attaching package: 'sjlabelled'

## The following object is masked from 'package:forcats':
##
##     as_factor

## The following object is masked from 'package:dplyr':
##
##     as_label

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

```
tab_model(lm1, lm2, lm3, lm4, lm5, lm6, lm7, lm8, lm9,
          title= "Table 1. Linear Panel Regression Models of Fixed Broadband
          Internet Access",
          dv.labels = c("lm1", "lm2", "lm3", "lm4", "lm5", "lm6",
                        "lm7", "lm8", "lm9"),
          string.p = "P-Value",
          show.ci = FALSE)
```

Table 1. Linear Panel Regression Models of Fixed Broadband Internet Access

lm1

lm2

lm3

lm4

lm5

lm6

lm7

lm8

lm9

Predictors

Estimates

P-Value

Estimates

P-Value

Estimates

P-Value

Estimates

P-Value

Estimates

P-Value

Estimates

P-Value

Estimates

P-Value

Estimates

P-Value

Estimates

P-Value

(Intercept)

0.05

<0.001

0.02

0.003

0.03

<0.001

0.04

<0.001

0.05

<0.001

0.06

<0.001

0.07

<0.001

0.05

<0.001

-0.02

0.001

gdpercap

0.00

<0.001

0.00

<0.001

0.00

<0.001

0.00

<0.001

0.00

<0.001

0.00

<0.001

0.00

<0.001

0.00

<0.001

0.00

<0.001

percapelect

-6.78

0.007

-1.71

0.491

percapmobsub

0.07

<0.001

Observations

1861

159

168

169

172

172

171

1850

1849

R2 / R2 adjusted

0.586 / 0.586

0.694 / 0.692

0.700 / 0.698

0.604 / 0.602

0.579 / 0.577

0.585 / 0.583

0.565 / 0.563

0.589 / 0.588

0.635 / 0.634

Reference

Alam, T. F., Sultana, N., & Rayhan, M. I. (2019). Structural equation modeling: An application of broadband penetration and GDP growth in Asia. Journal of Economic Structures, 8(1), 1–11. https://ideas.repec.org/a/spr/jecstr/v8y2019i1d10.1186_s40008-019-0148-0.html

Barry, J. (n.d.). COVID-19 exposes why access to the internet is a human right. OpenGlobalRights. https://www.openglobalrights.org/covid-19-exposes-why-access-to-internet-is-human-right/

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.7. https://CRAN.R-project.org/package=dplyr

International Telecommunication Union. (2017). ICT Facts and Figures 2017. https://about.fb.com/wp-content/uploads/2018/08/ICTFactsFigures2017.pdf

Khandelwal, Saloni & Agarwal, Sonali. (n.d.). Application Of Fixed Broadband Subscriptions Impact On Economic Growth. https://doi.org/10.21203/rs.3.rs-60231/v1

Lüdecke D (2018). "sjmisc: Data and Variable Transformation Functions." *Journal of Open Source Software*, *3*(26), 754. doi: 10.21105/joss.00754 (URL: https://doi.org/10.21105/joss.00754).

Lüdecke D (2021). *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.8.9, <URL: https://CRAN.R-project.org/package=sjPlot>.

Lüdecke D (2021). *sjlabelled: Labelled Data Utility Functions (Version 1.1.8)*. doi: 10.5281/zenodo.1249215 (URL: https://doi.org/10.5281/zenodo.1249215), <URL: https://CRAN.R-project.org/package=sjlabelled>.

Lüdecke, D. (2021, July 10). Summary of Regression Models as HTML Table. https://cran.r-project.org/web/packages/sjPlot/vignettes/tab_model_estimates.html

McDonough, Carol C. (2012, July 1). Fixed and mobile broadband: Demand and market structure. 23rd European Regional Conference of the International Telecommunications Society (ITS), Vienna, Austria. https://www.econstor.eu/bitstream/10419/60350/1/720272319.pdf

Minges, M. (2015). Digital Dividends: Exploring the Relationship Between Broadband and Economic Growth. https://openknowledge.worldbank.org/bitstream/handle/10986/23638/WDR16-BP-Exploring-the-Relationship-between-Broadband-and-Economic-Growth-Minges.pdf;sequence=1

Poushter, J. (2016, February 22). Internet access grows worldwide but remains higher in advanced economies. Pew Research Center. https://www.pewresearch.org/global/2016/02/22/internet-access-growing-worldwide-but-remains-higher-in-advanced-economies/

Revelle, W. (2021) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 2.1.6,.

Wempen, F. (2015). Different Types of Internet Connections—Dummies. https://www.dummies.com/education/internet-basics/different-types-of-internet-connections/

Wickham, Hadley & Grolemund, Garrett. (2017). R for Data Science. https://r4ds.had.co.nz/index.html

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

World Bank. (n.d). Fixed broadband subscriptions | Data Catalog. https://datacatalog.worldbank.org/fixed-broadband-subscriptions-1

Assembly, UN General. 2016. "Promotion and Protection of All Human Rights, Civil, Political, Economic, Social and Cultural Rights, Including the Right to Development." Koninklijke Brill NV. https://doi.org/10.1163/2210-7975_HRD-9970-2016149.

Brendan, Sasso. 2014. "In Net-Neutrality Push, Democrats Aim to Make the Internet a Utility." *The Atlantic*. https://www.theatlantic.com/politics/archive/2014/07/in-net-neutrality-push-democrats-aim-to-make-the-internet-a-utility/456897/.

Federal Communications Commission. 2014. "Types of Broadband Connections." *Federal Communications Commission*. https://www.fcc.gov/general/types-broadband-connections.

International Telecommunication Union. 2003. "The Birth of Broadband." https://www.itu.int/osg/spu/publications/birthofbroadband/faq.html.

———. 2016. "ICT Facts and Figutes 2016." https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf.

World Bank. n.d. "World Bank Open Data." *World Bank Open Data*. https://data.worldbank.org.