Final Project

Analysis of Internet Use in the World

Jiayi Ge, Yue Liu, Songqi Wang, Kexin Zhang, TUT 0112, L1

Introduction

- The Internet usage is an essential intermedia in globalization and modern society.
- The Internet usage is different around the world.
- How different are the Internet usage in different countries?
- What are the factors that have impact on Internet usage?

Objectives

- How can we define internet use in a country?
- Do different regions of the world have different internet usage?
- According to our definition of internet use, what is the impact of democracy, education, economy, and health on internet use?

Data

- Internet Usage = Internet Users/Population
- Democracy Score
- Education Expenditures
- GDP per capita
- Life Expectancy
- World Region

Data Cleaning

- Select the columns and combine them into one dataframe with proper names
- Change some character data, like GDP with $ sign, into type of double
- Correct one spelling mistake in the data of region
- In some small countries, more users than population. The IntUsage >1 are removed.

Data Summary

```
## Observations: 223
## Variables: 12
## $ Country         <chr> "Afghanistan", "Albania", "Algeria", "American...
## $ IntUsage        <dbl> 0.10349566, 0.66158944, 0.42205756, 0.33007145...
## $ Government_type <chr> "Authoritarian", "Hybrid regime", "Authoritari...
## $ DemocracyScore  <dbl> 2.55, 5.98, 3.56, NA, 3.62, NA, NA, 6.96, 4.11...
## $ Education_GDP   <dbl> NA, 3.3, 4.3, NA, 3.5, 2.8, 2.4, 6.3, 3.3, 6.0...
## $ GDP_Per_Capita  <dbl> 2000, 12500, 15200, 11200, 6800, 12200, 26300,...
## $ Lifetime        <dbl> 51.7, 78.5, 77.0, 73.4, 60.2, 81.5, 76.7, 77.3...
## $ Healthexpend    <dbl> 8.2, 5.9, 7.2, NA, 3.3, NA, 5.5, 4.8, 4.5, NA,...
## $ IntUsers        <dbl> 3531770, 2016516, 17291463, 17000, 2622403, 13...
## $ TeleLines       <dbl> 118769, 247010, 3130090, 10000, 161070, 6000, ...
## $ Population      <dbl> 34124811, 3047987, 40969443, 51504, 29310273, ...
## $ Region          <chr> "Asia & Pacific", "Europe", "Arab States", "As...
```

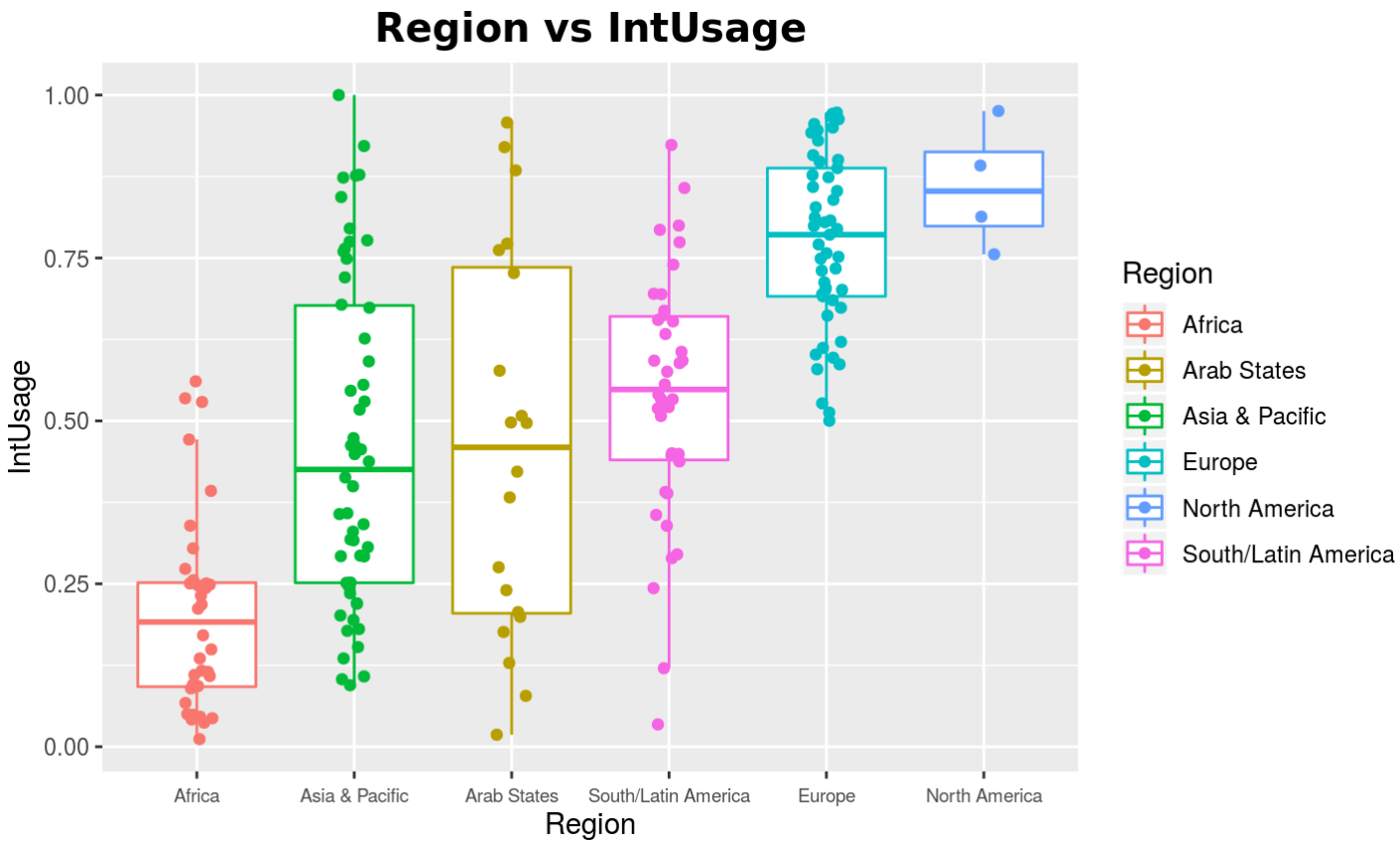
Statistic method 1:Boxplot

- We create boxplot by using geom_boxplot() fuction in R language which belongs to library(ggplot2) package to describe the relationship between the Region and IntUsage.
- We create boxplot by using geom_boxplot() fuction in R language which belongs to library(ggplot2) package to describe the relationship between the Government_type and IntUsage.

Statistic method 2: linear regression

- We create the linear regression by using lm() fuction to describe the relationship between IntUsage and Region.
- We create the linear regression by using lm() fuction to describe the relationship between IntUsage and DemocracyScore.
- We create the linear regression by using lm() fuction to describe the relationship between IntUsage and Education_GDP.
- We create the linear regression by using lm() fuction to describe the relationship between IntUsage and GDP_Per_Capita.
- We create the linear regression by using lm() fuction to describe the relationship between IntUsage and Lifetime.

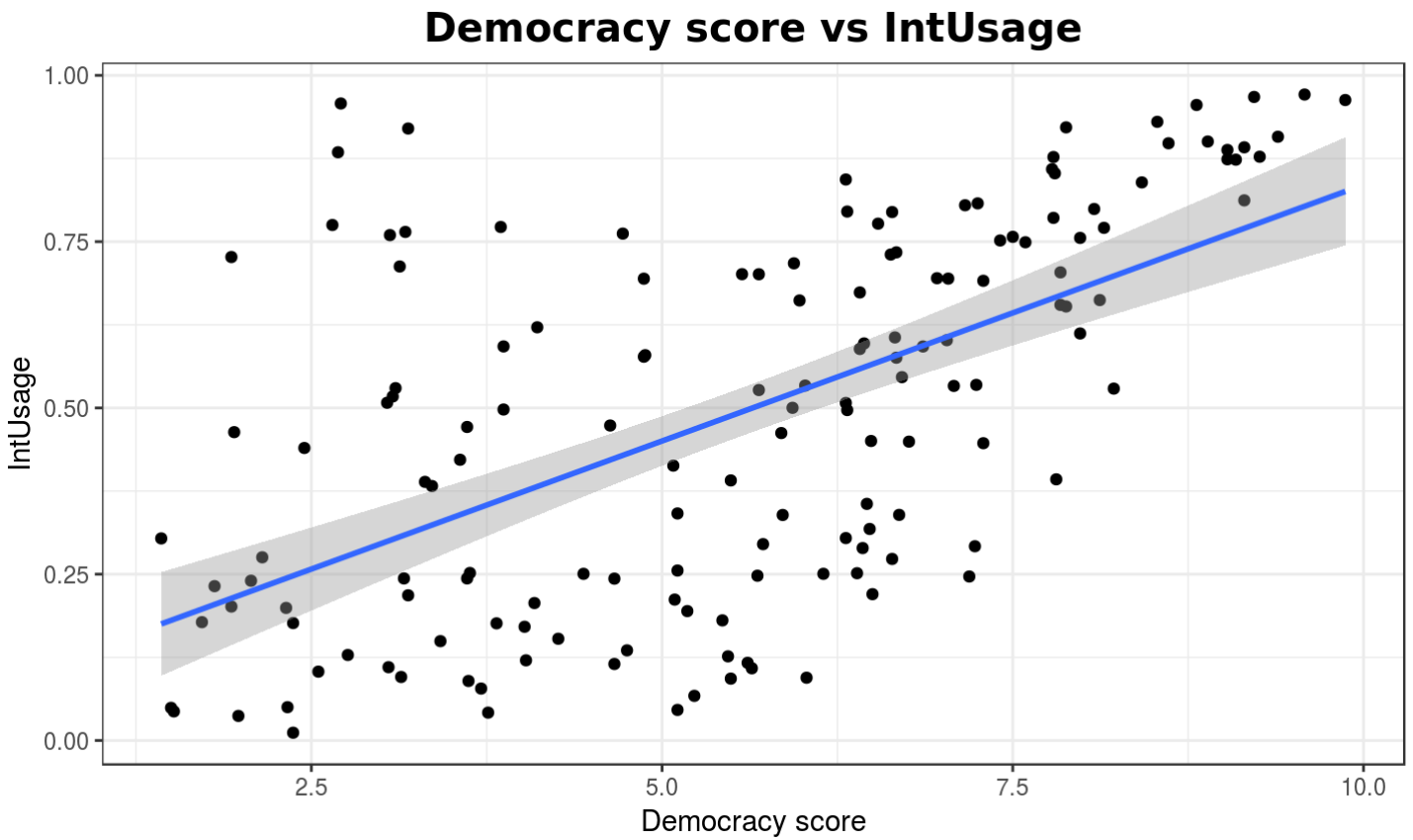
Region


Region vs IntUsage
1.00
0.75
0.50
0.25
0.00
IntUsage
Africa
Asia & Pacific
Arab States
South/Latin America
Europe
North America
Region
Region
Africa
Arab States
Asia & Pacific
Europe
North America
South/Latin America

Region

```
##                            Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)                0.462700    0.02775 16.67000 2.264e-39
## RegionAfrica              -0.258900    0.04388 -5.89900 1.591e-08
## RegionArab States         -0.001206    0.05338 -0.02259 9.820e-01
## RegionEurope               0.312500    0.04024  7.76600 4.500e-13
## RegionNorth America        0.396400    0.10570  3.75100 2.325e-04
## RegionSouth/Latin America  0.075650    0.04318  1.75200 8.136e-02
```

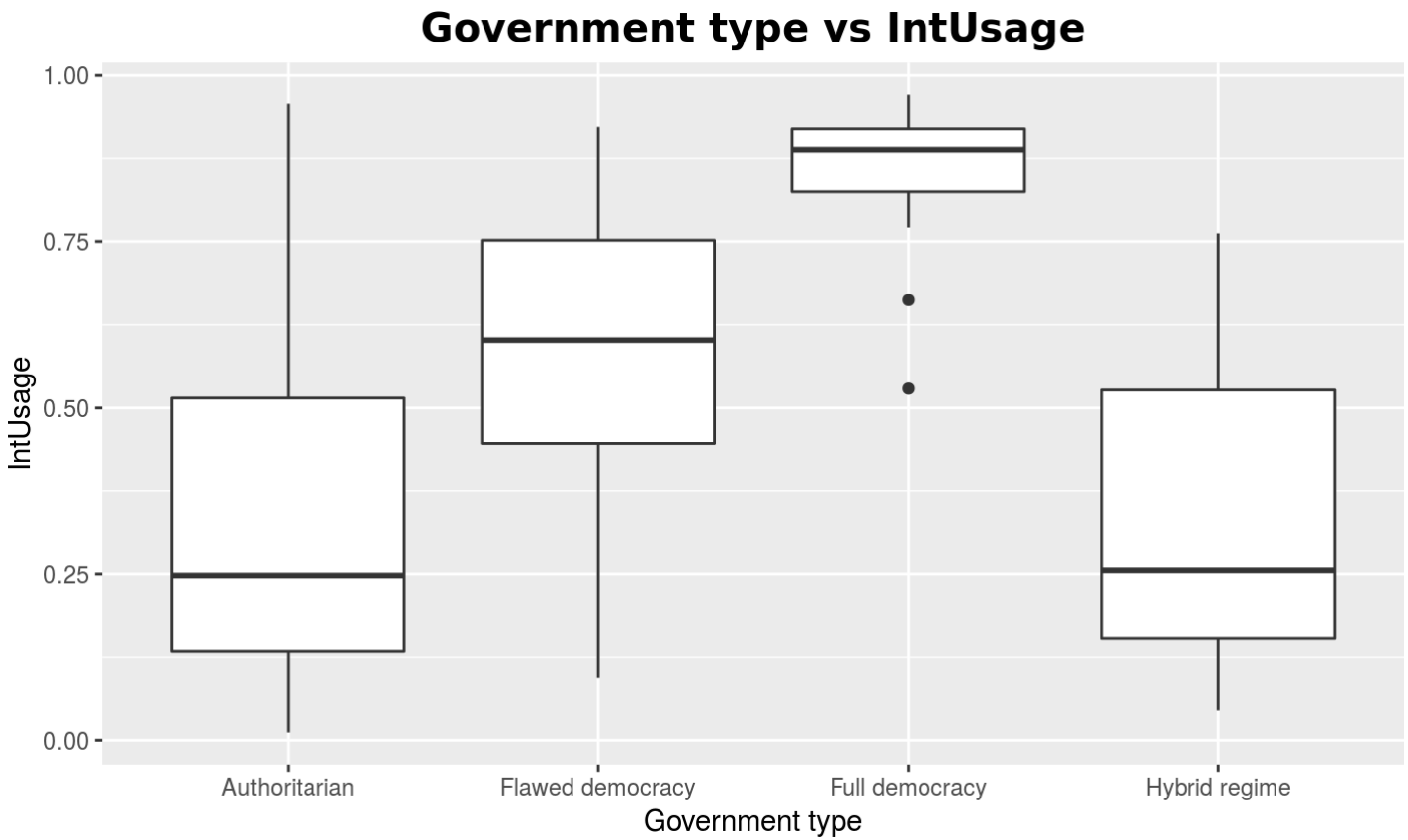
Democracy


Democracy score vs IntUsage
1.00
0.75
0.50
0.25
0.00
IntUsage
2.5
5.0
7.5
10.0
Democracy score

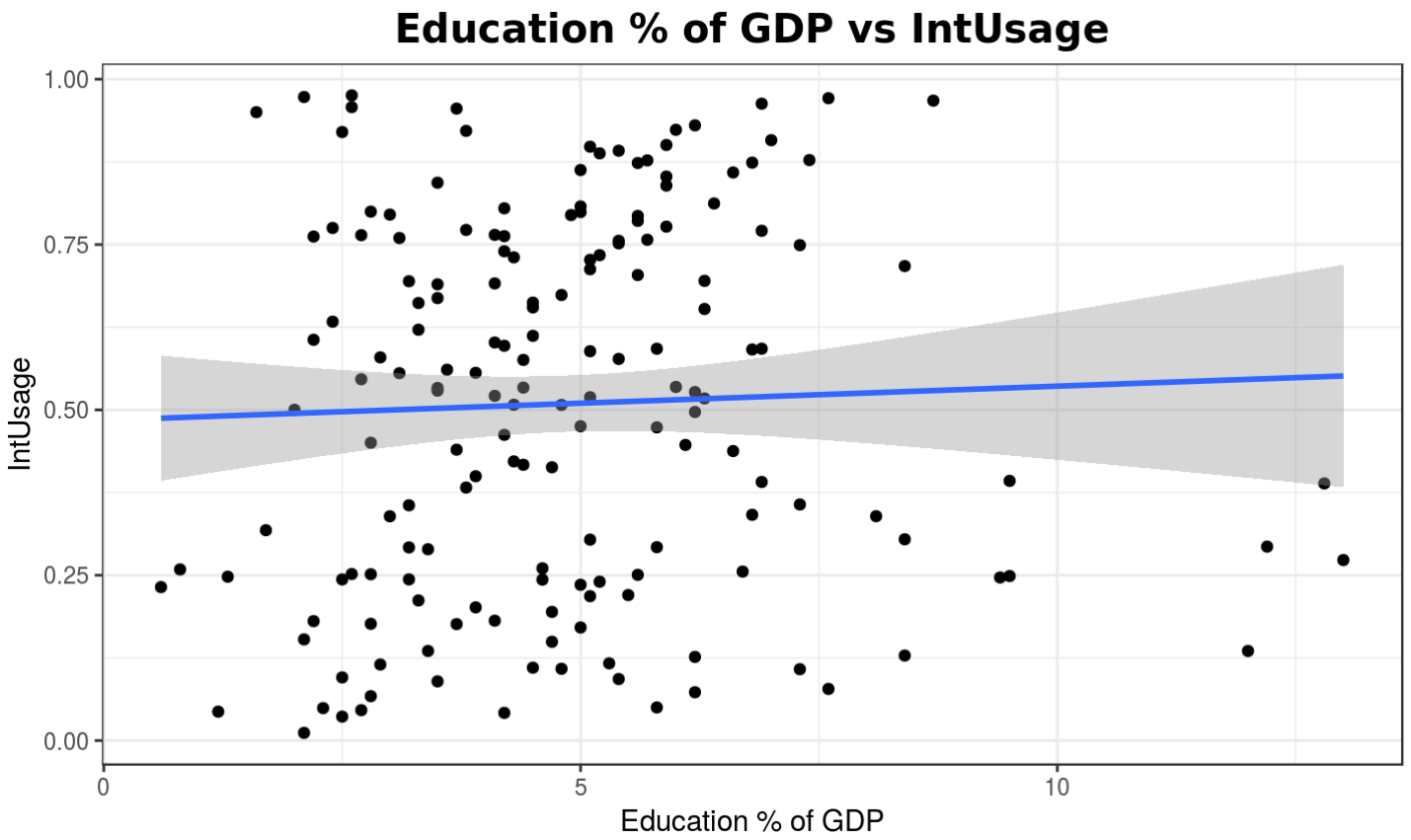
Democracy

```
##
## Call:
## lm(formula = IntUsage ~ DemocracyScore, data = compare_Dem_usage)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.43530 -0.18077  0.00475  0.14439  0.68386
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.065091   0.050490   1.289    0.199
## DemocracyScore 0.077045   0.008501   9.063 5.68e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.228 on 153 degrees of freedom
## Multiple R-squared:  0.3493, Adjusted R-squared:  0.3451
## F-statistic: 82.14 on 1 and 153 DF,  p-value: 5.676e-16
```

Democracy


Government type vs IntUsage
1.00
0.75
0.50
0.25
0.00
IntUsage
Authoritarian
Flawed democracy
Full democracy
Hybrid regime
Government type

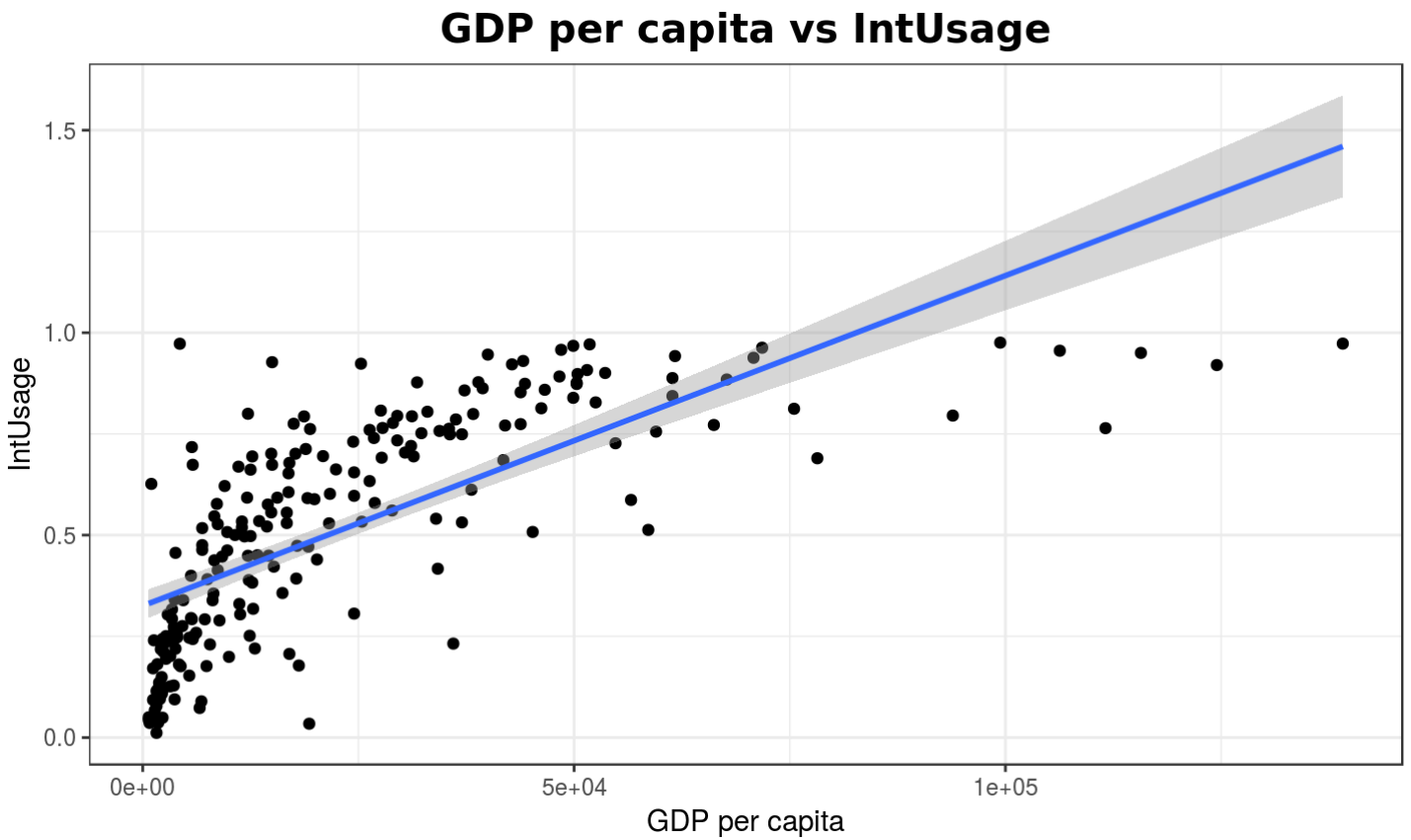
Education


Education % of GDP vs IntUsage
IntUsage
1.00
0.75
0.50
0.25
0.00
0
5
10
Education % of GDP

Education

```
##
## Call:
## lm(formula = IntUsage ~ Education_GDP, data = compare_edu_usage)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48350 -0.25630  0.01942  0.25379  0.47790
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.48436    0.05341   9.068 2.98e-16 ***
## Education_GDP  0.00515    0.01010   0.510    0.611
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2817 on 169 degrees of freedom
## Multiple R-squared:  0.001535,   Adjusted R-squared:  -0.004373
## F-statistic: 0.2598 on 1 and 169 DF,  p-value: 0.6109
```

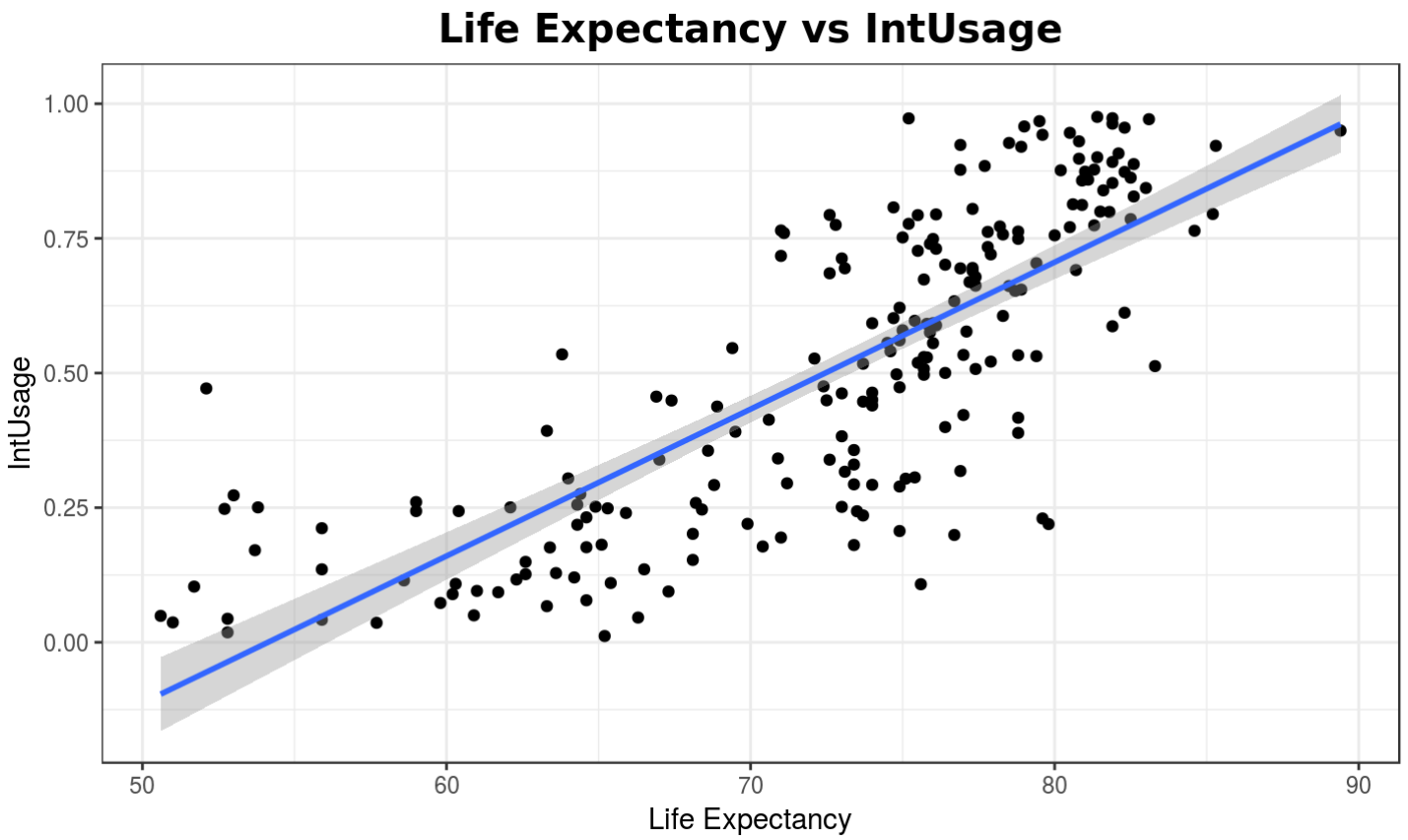
Economy


GDP per capita vs IntUsage
IntUsage
1.5
1.0
0.5
0.0
0e+00
5e+04
1e+05
GDP per capita

Economy

```
##
## Call:
## lm(formula = IntUsage ~ GDP_Per_Capita, data = compare_gdp_usage)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48686 -0.15150  0.01641  0.14950  0.61226
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.254e-01  1.820e-02   17.88   <2e-16 ***
## GDP_Per_Capita 8.156e-06  5.396e-07   15.12   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1961 on 217 degrees of freedom
## Multiple R-squared:  0.5129, Adjusted R-squared:  0.5106
## F-statistic: 228.5 on 1 and 217 DF,  p-value: < 2.2e-16
```

Health


Life Expectancy vs IntUsage
IntUsage
1.00
0.75
0.50
0.25
0.00
50
60
70
80
90
Life Expectancy

Health

```
##
## Call:
## lm(formula = IntUsage ~ Lifetime, data = compare_life_usage)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.48082 -0.10786 -0.00331  0.12509  0.52664
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.476503   0.107994  -13.67   <2e-16 ***
## Lifetime     0.027279   0.001474   18.50   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1744 on 213 degrees of freedom
## Multiple R-squared:  0.6165, Adjusted R-squared:  0.6147
## F-statistic: 342.4 on 1 and 213 DF,  p-value: < 2.2e-16
```

Final Model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.002896e+00	1.274508e-01	-7.868886e+00	1.179632e-12
DemocracyScore	1.704953e-02	6.626141e-03	2.573071e+00	1.119362e-02
GDP_Per_Capita	4.794072e-06	6.680423e-07	7.176300e+00	4.763440e-11
Lifetime	1.750179e-02	2.005383e-03	8.727403e+00	1.038720e-14
Education_GDP	8.229308e-03	5.733634e-03	1.435269e+00	1.535937e-01

R-squared:

```
## [1] 0.8078404
```

Conclusion

- We do have a difference between the internet usages of different countries. Generally, developed countries or developed regions have a higher Internet usage.
- Economy and health have a strong positive impact on internet usage.
- Democracy level have a week positive impact on internet usage.
- Education has no impact on internet usage.

Acknowledgment

We are thankful to our Prof.Taback for his knowledge that provided to us during the course of this seminar and his presentation.

Especially thanks to our TA, Yang Zhu for his valuable guidance and encouragement.

Last but not least, we would also want to extend our appreciation to those could not be mentioned here due to the limitation of time.