

# Analysis of Trends between GDP and Infrastructure of Countries

160175125

## Introduction

The aim of the analysis was to see whether any relationship exists between the wealth of a country and its infrastructure characteristics. More specifically, what are the major differences between high GDP per capita countries and low GDP per capita countries. The data used was taken from The American Central Intelligence Agency “World Factbook”[1].

## Exploratory Analysis

The dataframe “Infrastructure”, created for the analysis (See Appendix.2), contains the data for each country’s wealth, communications, traffic, and population density, so has columns:

- Country,
- GDP : GDP per capita,
- Landlines : The average number of landlines per person,
- Mobiles : The average number of mobile phone subscriptions per person,
- Internet : The average number of internet connections per person,
- Roads : The average number of people sharing a 1 kilometre stretch of road,
- Airports : The average number of people per airport,
- Pop Density : The average number of people per square kilometre of country.

Broadband subscriptions from the original dataset (See Appendix.1) were omitted from the analysis as anyone with a broadband subscription must have an internet connection by default. Therefore the internet data was sufficient to capture the information needed for this particular analysis.

The summary of the Infrastructure dataset (See Appendix.3) shows the summerized data.

The minimum value of Landlines (Guinea) was checked for validity as it looked suspicious, possibly due to a rounding error. The original dataset also states that there are no landlines in Guinea (See Appendix.5). This is certainly a mistake however since Export.gov[2] states that landlines do exist, although not preferred over mobile phones in Guinea. This data point was omitted to prevent complications with log transformations used to create the plots.

A PCA analysis was conducted with the data “Landines”, “Mobiles”, “Internet”, “Roads”, “Airports”, and “Pop Density”. The summary of the analysis (See Appendix.4) states that the first 4 principle components explain over 90% of the variation in the data, which was deemed sufficient for the analysis. Therefore, the 5th and 6th were omitted from the analysis.

Plot 1 is a scree plot illustrating the relative importance of each principle component. As can be seen PC5 and PC6 do not account for much of the variation.

In table 1, the principle components are shown as a composition of the original catagories. Following is the interpretation of the infrastructure data using these principle components.

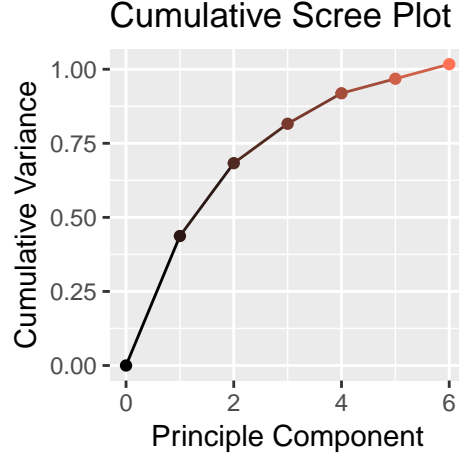


Figure 1: Plot 1: Infrastructure Scree Plot

Table 1: Principle Components Composition

	PC1	PC2	PC3	PC4
Landlines	-0.4972540	-0.2536141	-0.0221695	0.5395808
Mobiles	-0.3872631	-0.2799967	-0.4080555	-0.7492711
Internet	-0.5336360	-0.2572103	-0.1450380	0.2438873
Roads	0.4227186	-0.4801064	-0.1771841	0.0729267
Airports	0.3723848	-0.4916952	-0.4075495	0.1936183
Pop Density	-0.0252358	-0.5646753	0.7838829	-0.2125044

Figure 1 shows every combination of the first 4 principle components as 2 dimensional plots. Each data point (Country) is size and colour coded to represent it's GDP per capita. Countries with a higher GDP per capita are lighter coloured and are larger in size. The red hue plots have PC1 along the x-axis, the blue hue plots have PC2 along the x-axis, and the yellow hue plot has PC3 along the x-axis. The legend in each plot shows the minimum, maximum, median, and 1st and 3rd quartiles of the GDP per capita (GDPpp). A small amount of outliers are missing in the plots due to the choice of scaling, but were not important to the analysis. In each plot there is a clear split between the infrastructure characteristics of high GDP per capita countries and low GDP per capita countries. The discussion is also in terms of generalities, so a mention of "all" countries refers to the vast majority not including outliers.

- Plot A1 shows a slight positive linear relationship between high GDP per capita countries and low GDP per capita countries, and also a positive relationship between PC1 and the variance of PC2. The higher GDP countries are found in the bottom left of the plot, while the lower GDP countries are spread around the center right.

The concentration of high GDP countries in the bottom left implies that high GDP countries have a good communications structure and high population density, and the plot says little about the transport networks. If amount of people in comparison with the transport networks had a substantial effect on the values of PC1 and PC2 then we would expect to see a negative linear relationship between PC1 and PC2. As we do not see this, it suggests that transport networks do not play a big part in the plot.

Low GDP countries have much poorer communications infrastructure in comparison to the high GDP countries, and therefore must vary more in terms of population density, which explains the increase in variance of PC2 as PC1 increases. The population density is the driver for the slope in the data.

- Plot A2 shows that as PC1 increases, the variance of PC3 increases slightly, and no other relationship exists.

PC3 implies that for higher GDP countries, there is a large amount of people relative to the transport networks of those countries. The implication exists because higher GDP countries always have good communications structures and high population density. This characteristic is also very similar across nearly all high GDP countries due to the low variance in the vertical direction.

For low GDP countries, given that they generally have poor communications, if the country has a high population density, then it will also have a large amount of people in comparison with its transport networks. Also, if a country has a low population density, then there will be less people in comparison with its transport networks. Low GDP countries with higher PC3 values should have less problems with its transport networks in comparison to its population than those further down with a similar transport network. This explains the variance seen in PC3.

- Plot A3 shows a large negative linear relationship between PC1 and the variance of PC4.

For high GDP countries, PC4 implies there is a large variation between countries that depend more on fixed communication, and those that depend more on wireless communication. This implication is due to high GDP countries always having a high population and high number of people in comparison to its transport networks, of which the effects cancel out in PC4. Also, the highest GDP countries tend to favour a balance between fixed and wireless communication, or favour fixed communication overall.

For lower GDP countries, the variation becomes much less, simply because they always have poor communication structures overall. This explains the variance of PC4.

- Plot A4 compares the population density against congestion problems. As the population density increases, so do the congestion problems.
- Plot A5 compares the population density against whether a country relies on mainly a fixed communications structure, or a wireless communications structure. The plot shows no relation between these variables.
- Plot A6 compares congestion problems against whether a country relies on mainly a fixed communications structure, or a wireless communications structure. The plot suggests that countries that favour fixed communications or a balance between fixed and wireless communication tend to have the lowest congestion.

## Conclusion

Overall, the analysis has highlighted some interesting differences in the characteristics of a country's infrastructure, depending on the relative wealth of that country.

High GDP per capita countries are similar in population density, quality of communication structures, and magnitude of congestion problems. However they vary a lot in the type of communication structure, either fixed, or wireless dominant. Also, fixed communication dominant countries seem to perform slightly better than wireless dominant countries in terms of GDP per capita and quality of communications structure.

Low GDP per capita countries are more varied in population density and in the magnitude of congestion problems, however they are similar in their communications type, and quality, most likely due to having poor communication structures altogether.

## References

- [1] The Central Intelligence Agency. "The World FactBook". <https://www.cia.gov/library/publications/the-world-factbook/>. Accessed October 2018.
- [2] Export.gov. "Guinea - Telecommunications/Electric". <https://www.export.gov/article?id=Guinea-Telecommunications>. Accessed October 2018.

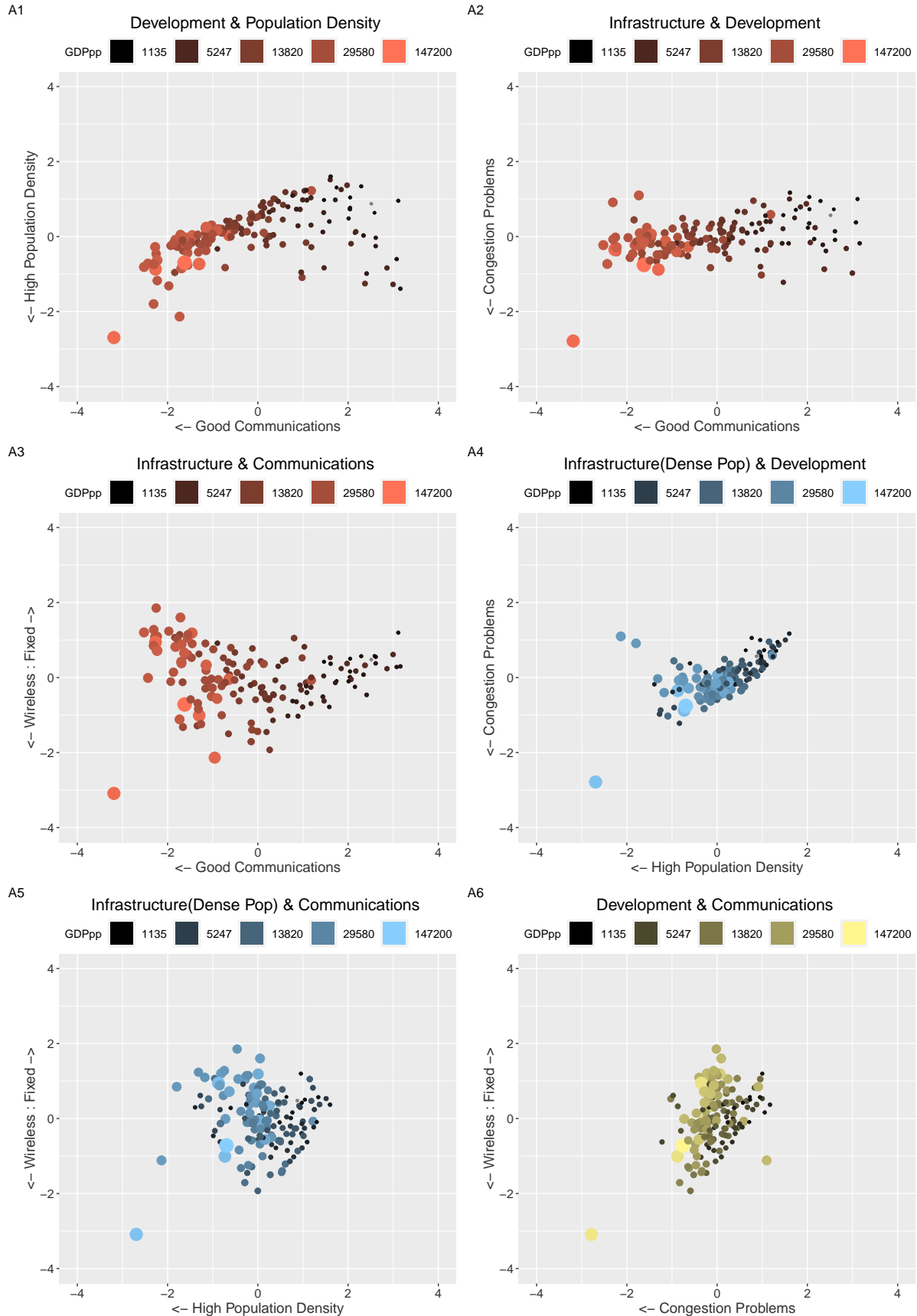


Figure 2: Figure 1: Principle Component Plots

## Appendix

### Appendix.1 (Countries dataset sample)

```
##          country      area population popgrwth  age infmort lifeexp
## 1      Afghanistan  652230   34124811    2.36 18.8   110.6   51.7
## 2          Albania   28748    3047987    0.31 32.9    11.9   78.5
## 3          Algeria  2381741   40969443    1.70 28.1    19.6   77.0
## 4          Angola  1246700   29310273    3.52 15.9    67.6   60.2
## 5 Antigua and Barbuda    443     94731    1.21 31.9    12.1   76.7
## 6          Argentina 2780400   44293293    0.91 31.7     9.8   77.3
##  obesity      gdp hlthexp unemp      co2  phone  mobile internet
## 1      5.5 6.955e+10    8.2  23.9 7.40e+06 118769 23929713 3531770
## 2     21.7 3.597e+10    5.9  14.0 4.30e+06 247010 3497950 2016516
## 3     27.4 6.329e+11    7.2  11.7 1.28e+08 3130090 49873389 17291463
## 4      8.2 1.903e+11    3.3   6.6 3.30e+07 161070 13323952 2622403
## 5     18.9 2.393e+09    5.5  11.0 6.00e+05  22504  180000   60000
## 6     28.3 9.202e+11    4.8   8.1 4.32e+07 9530349 61897379 30786889
## broadband airports roadways
## 1      16810      52   42150
## 2     293623      4   18000
## 3    3166907     157  113655
## 4     96919     176   51429
## 5      9261      3   1170
## 6    7870222    1138  231374
```

### Appendix.2 (Infrastructure dataset sample)

```
##          Country      GDP      Landlines  Mobiles  Internet  Roads
## 1      Afghanistan  2038.107 0.003480430 0.7012409 0.10349566 809.60406
## 2          Albania 11801.231 0.081040372 1.1476263 0.66158944 169.33261
## 3          Algeria 15448.099 0.076400599 1.2173314 0.42205756 360.47198
## 4          Angola  6492.604 0.005495343 0.4545830 0.08947044 569.91723
## 5 Antigua and Barbuda 25261.002 0.237556872 1.9001172 0.63337239 80.96667
## 6          Argentina 20775.154 0.215164607 1.3974436 0.69506887 191.43591
##  Airports Pop Density
## 1 656246.37    52.32021
## 2 761996.75   106.02431
## 3 260951.87    17.20147
## 4 166535.64    23.51029
## 5 31577.00    213.83973
## 6 38922.05     15.93055
```

### Appendix.3 (Infrastructure dataset summary)

```
##          Country      GDP      Landlines
## Afghanistan      : 1  Min.      : 1135  Min.      :0.00000
## Albania           : 1  1st Qu.: 5247  1st Qu.:0.02802
## Algeria           : 1  Median : 13817 Median :0.14027
## Angola            : 1  Mean    : 22115 Mean    :0.16598
## Antigua and Barbuda: 1  3rd Qu.: 29580 3rd Qu.:0.25200
## Argentina         : 1  Max.    :147171 Max.    :0.57713
```

```

## (Other) :152
##      Mobiles      Internet      Roads      Airports
## Min.   :0.08549   Min.   :0.01167   Min.   : 12.15   Min.   : 3539
## 1st Qu.:0.89419   1st Qu.:0.25078   1st Qu.: 122.47   1st Qu.: 63293
## Median :1.13546   Median :0.52799   Median : 218.98   Median : 179045
## Mean   :1.11702   Mean   :0.50814   Mean   : 451.49   Mean   : 444936
## 3rd Qu.:1.30638   3rd Qu.:0.75467   3rd Qu.: 473.11   3rd Qu.: 480877
## Max.   :3.26493   Max.   :0.97121   Max.   :7420.50   Max.   :8768143
##
##      Pop Density
## Min.   : 1.962
## 1st Qu.: 29.238
## Median : 75.975
## Mean   : 190.651
## 3rd Qu.: 141.556
## Max.   :8448.961
##

```

#### Appendix.4 (Infrastructure PCA summary)

```

## Importance of components:
##              Comp.1   Comp.2   Comp.3   Comp.4   Comp.5
## Standard deviation 1.6206937 1.2158463 0.8954711 0.7879347 0.54552689
## Proportion of Variance 0.4377747 0.2463804 0.1336448 0.1034735 0.04959993
## Cumulative Proportion 0.4377747 0.6841551 0.8177998 0.9212733 0.97087326
##
##              Comp.6
## Standard deviation 0.41804357
## Proportion of Variance 0.02912674
## Cumulative Proportion 1.00000000

```

#### Appendix.5 (Guinea data)

```

##      Country      GDP Landines      Mobiles      Internet      Roads Airports
## 61  Guinea 2132.293      0 0.8699948 0.09546969 279.9194 775866.7
##      Pop Density
## 61      50.49223

```