Worldbank Internet Usage Report

Tom Wheetley

10-26-2022

Contents

1 Introduction

This is the "Choose Your Own, (CYO), project for the Harvard X data Science CapStone Course. I will be looking at internet usage and how poverty can effect access to it as well as how decling birth rates are often see in countries that are improving their economic environments and this can lead to more internet access.

2 Overview

Birth rate in general, decline as the economic picture improves and we can see internet usage increases in these countries. The World bank in 2017 set the poverty line to $2.15 cents per person, per day. I will be using this mark to try to predict when a country may cross that threshold.

3 Executive Summary

During the COVID-19 pandemic, many health agencies and service providers were forced to close their offices to prevent transmission of the virus. However to serve their clients many providers switched to providing medical care through tele-health services. The providers were not able to provide all services over the phone, some required internet access. For many developing countries access to the internet was a factor in gaining access to critical care. Many families dealing with poverty and the economic consequences of the pandemic were unable to access the care or information they required. In addition, populations with worse internet access also tend to have higher rates of chronic conditions and worse health outcomes, suggesting that as a country develops the reliance on internet becomes even more critical. This does not suggest that internet access pull people out of poverty, but rather is the result of improving economic conditions.

#Key components: Function RMSE -Root Mean Square Error (RMSE) will be used to gauge the success of the predictions, measuring the difference between the actual and predicted value. Two different models will be used to evaluate RMSE, A more accurate prediction will be identified by a lower RMSE.

The following is the calculations for RMSE:

$$RMSE = \sqrt{\frac{1}{N}\sum_{u,i}(\hat{y}_{u,i} - y_{u,i})^2}$$

4 Data downloads - There are 5 data sources for the project 3 are

downloaded automatically, the internetusage.csv file and the poverty_explorer.csv are included with the uploaded files and will have to be loaded. The code is provided and the “Choose” function is used to browse to the file. ## Data cleansing - Cleansing of the data is performed on the data sets removing nulls and NAs removing and renaming columns for ease of use ## Visualizing the data - Data will be visualized with 4 plots. The first showing birth Rates and percent of internet users by income group. The second shows Birth Rate by Region and the percentage of people with internet access. The Third shows mean internet usage by year, by region The forth shows the mean poverty by region for the same time frame as plot 3 for the same regions

5 Data visualizations

The poverty, internet usage and birth rate data are combined and loaded into two data sets: pov_exp for training purpose and valid for the validation.

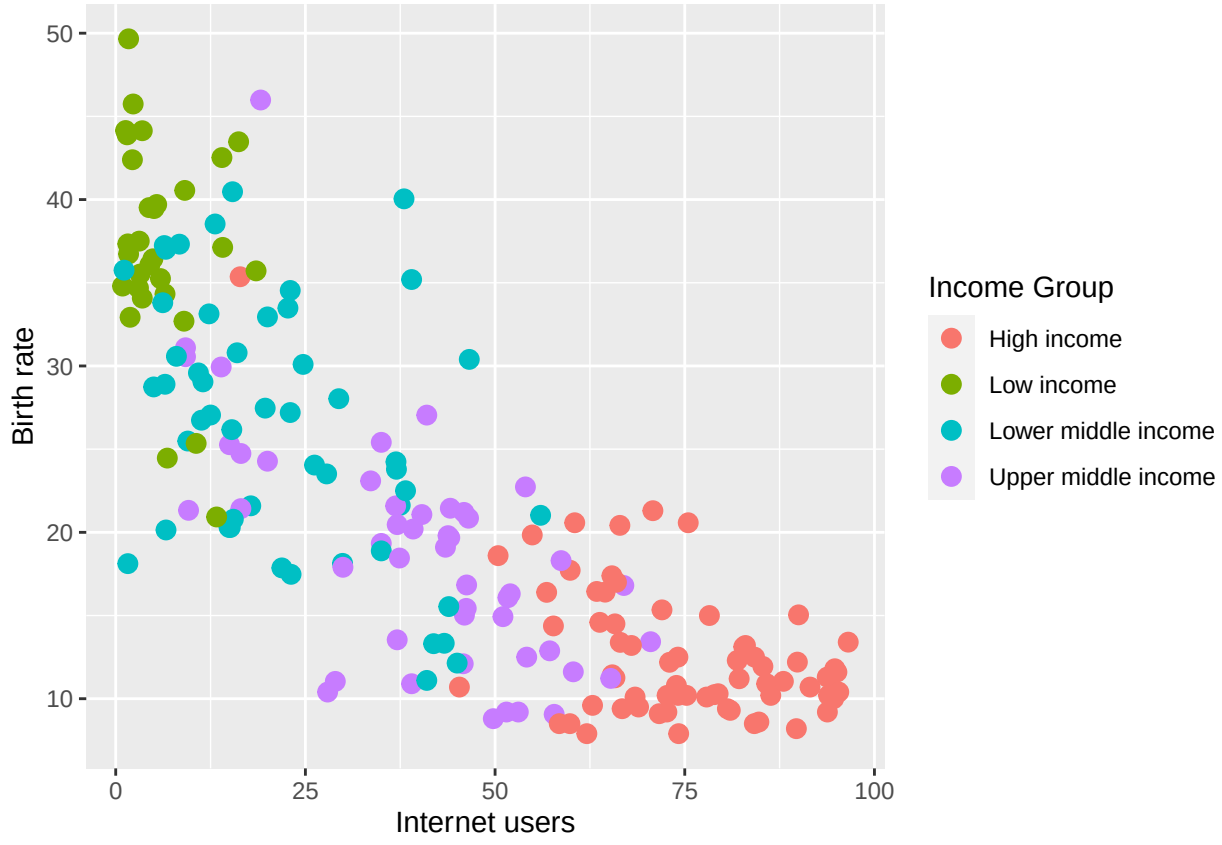
There are total of 3 columns in both pov_exp and valid datasets: pct_below is data type num shows the percentage of people with in a coutry living below th poverty line country is data type chr and give the name of the country Year is the year for the sample

#Birth Rates and Internet Usage

##Birth rate is defined as the number of births per 1000 adults. The plot shows the relationship between declining birth rates and the percentage of the population with internet access.The higher incomees show a greater internet usage and lower birth rate.

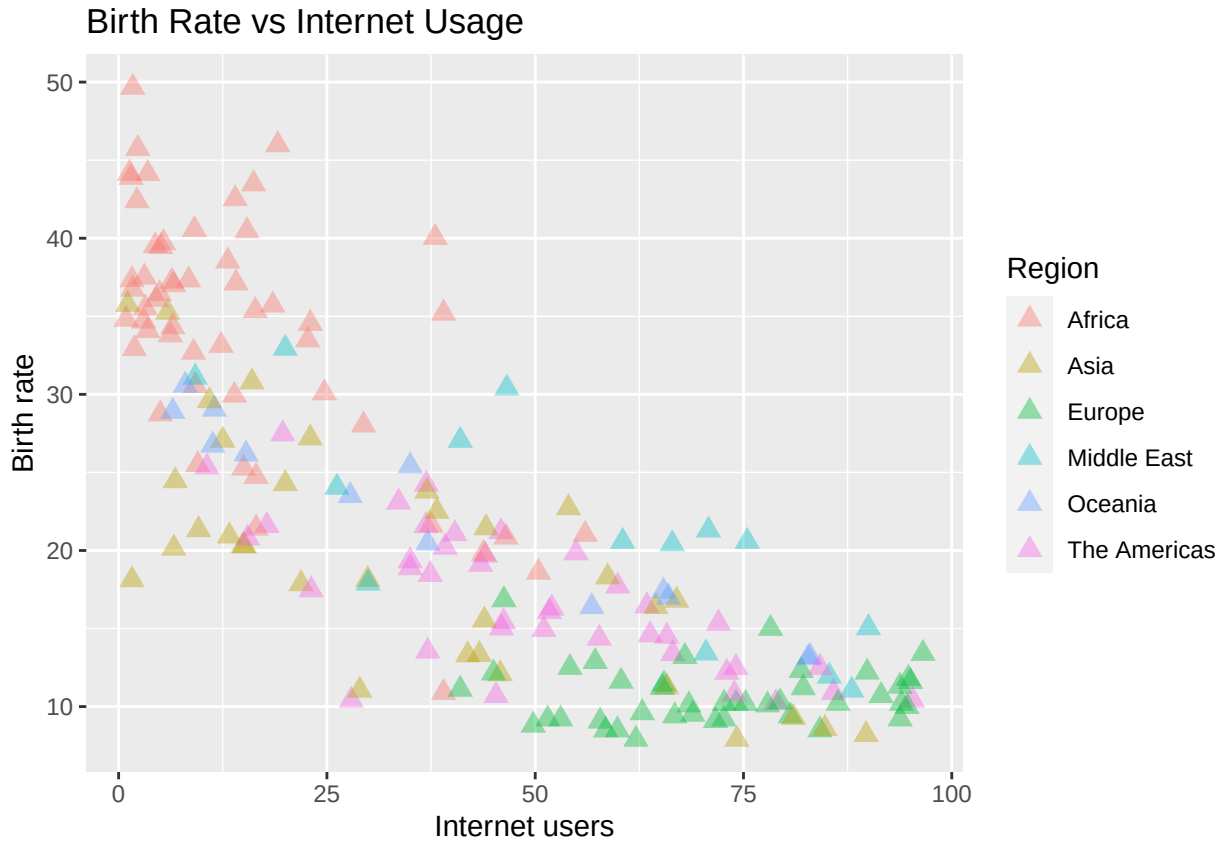
5.1 PLOT ONE

PLOT ONE - Birth Rates and internet use by income group You can see the groups with the lower birth rate are generally in the higher income group and have a higher higher percentage of internet adoption.



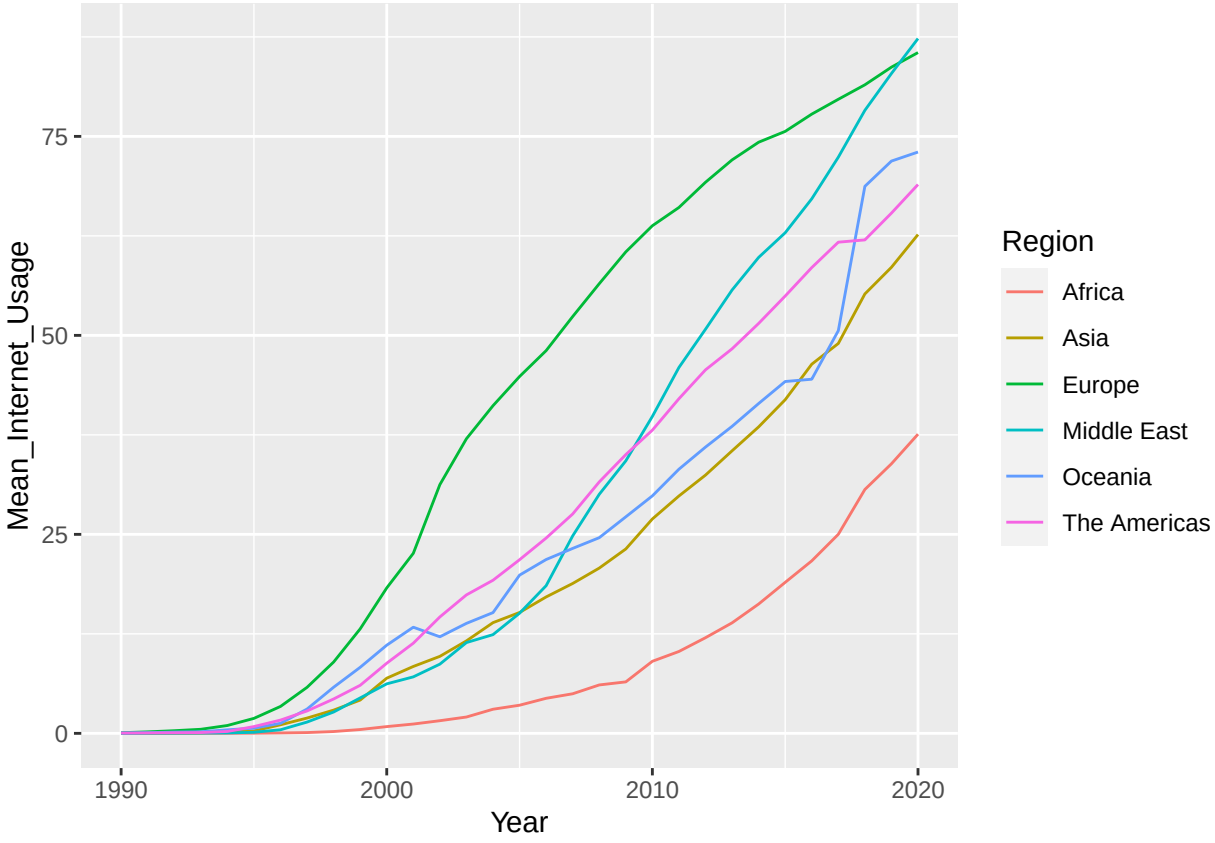
5.2 PLOT TWO

PLOT TWO - Birth Rates vs Internet usage by Continent. (Africa shows the slowest adoption) In the plot below you can see the Africa triangles are predominately in the upper left showing a higher birth rate than the higher income countries of Europe represented by the green triangles.



5.3 PLOT THREE

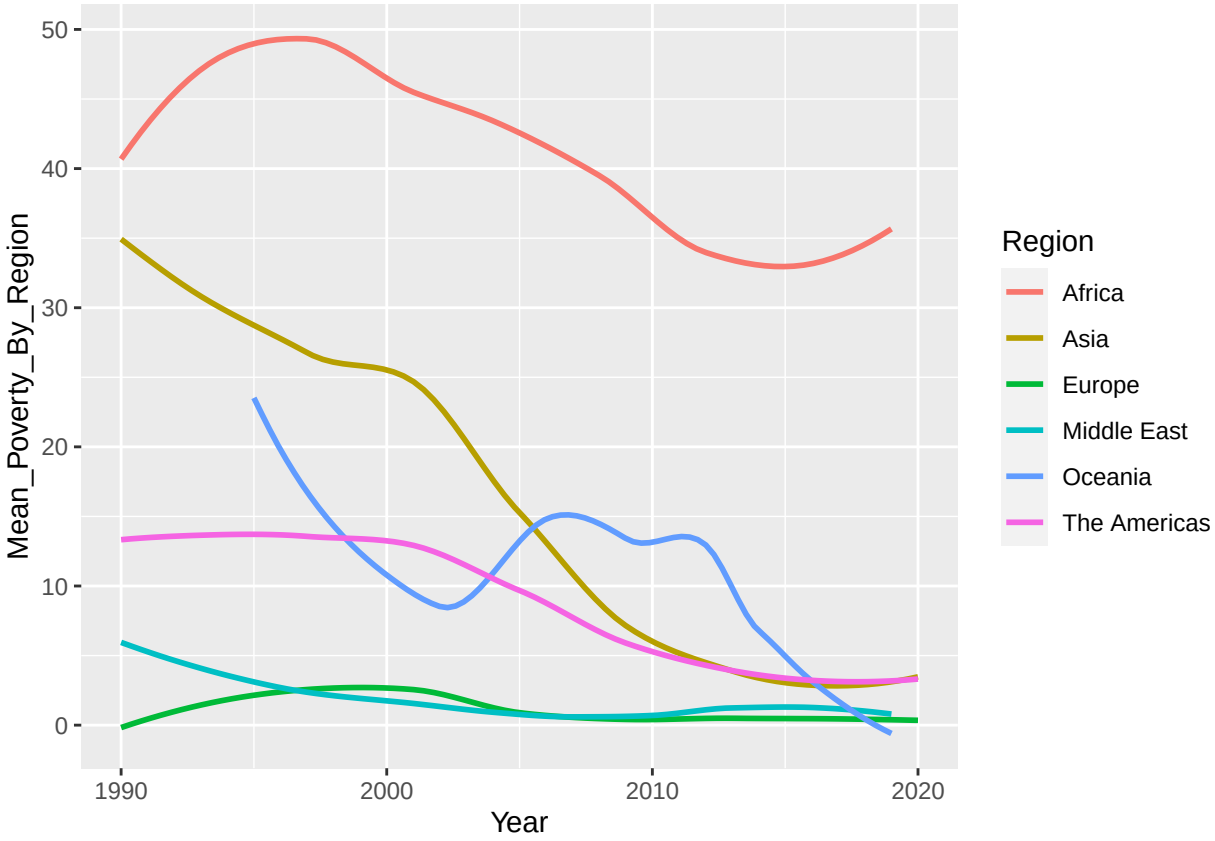
PLOT THREE - we can see the adoption of the internet amoung the develop countries of teh world well over 50% by 2020, however Africa lags behind around 37%.




```
## [1] FALSE
```

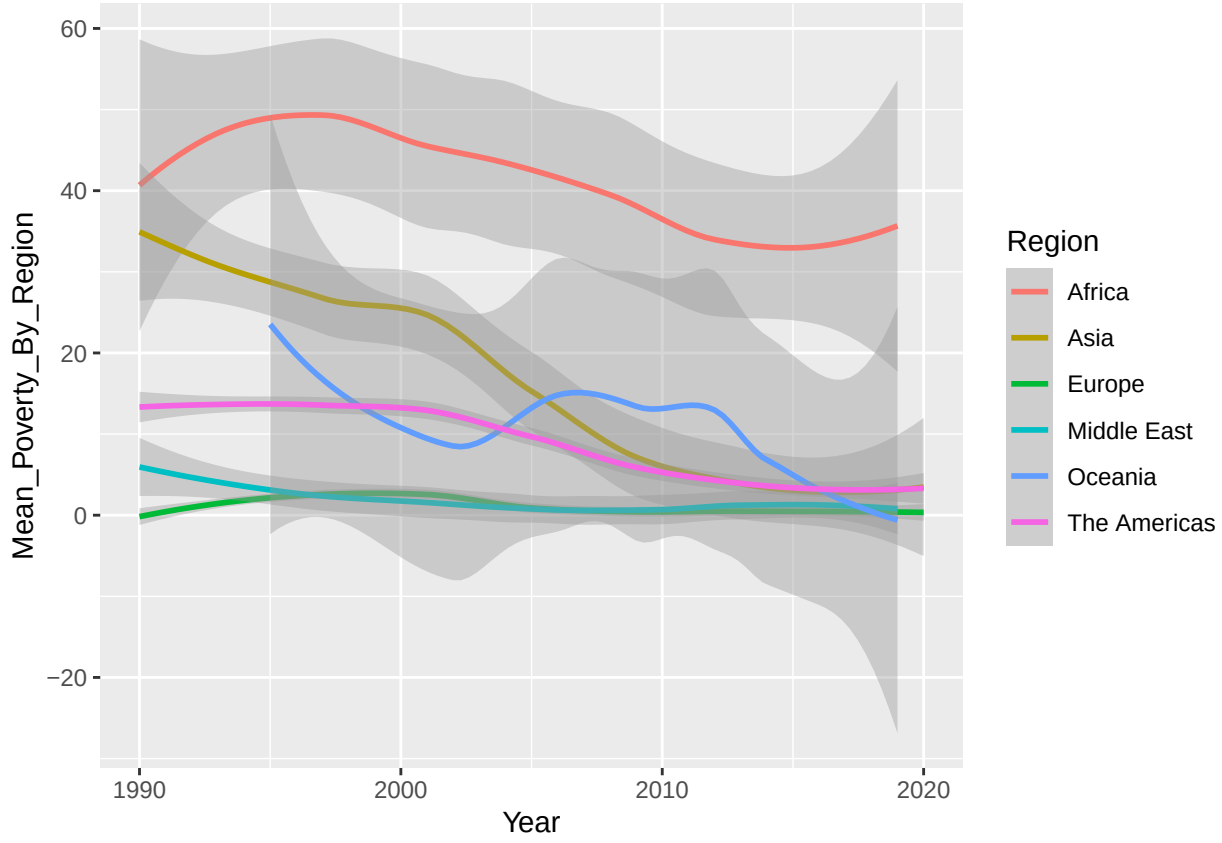
5.4 PLOT FOUR

PLOT FOUR - show us the mean poverty by region over the last 30 years.



5.5 PLOT FIVE

PLOT FIVE - shows the decline in poverty by region and the years, and can be compared to the previous plot to show the decline in poverty and the increase in internet usage. As the countries develop, birth rates decline and adoption of technology increases.



print(results)

```
## [1] -10.91163
```

6 Results

Results Section: The mean-baseline model appears to be the more accurate of the two.

```
##                 model     RMSE
## 1 Mean-Baseline Model 17.09820
## 2 Poverty Based Model 27.39214
```

print(results)

7 Conclusion

In conclusion the data shows that people in countries climbing out of poverty are in a better economic position to access the internet.This bring other benefits like better access to information and healthcare. The small data sets make accurate reporting difficult, but overall you can see the connection to improved economic situations and the adoption of the internet.

This model could likely be further improved by including other parameters (time of day,time of year,year, genre), or by using more sophisticated machine learning algorithms (knn, neural network, etc.).