Employment rates and Internet Access

Name: Sean, Alex, Yike Li (Lynnie)
Professor: Ott toomet
Section:AG(Group3)
Info201

TABLE OF CONTENTS

01

INTRODUCTION

Background

RESEARCH QUESTIONS

Is COVID-19 the reason for internet use to rise or was it mainly due to remote working?

ALEX

Does internet access show any significant increase in jobs for under-developed countries?

SEAN

As internet is more incorporated into our lives, will it create more jobs and increase employment rates in developed countries?

Lynnie

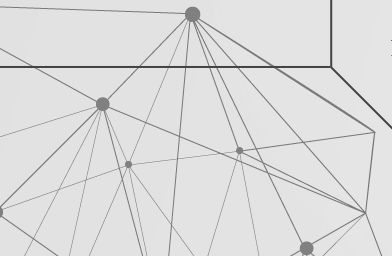
Why this topic & What we know

COVID-19 affected everyone, in both the illness itself, and the after-effects of (in this case) increased internet usage and employment rates.

By 2019, about 57% of people in the world was using the internet.

While the internet has grown with many of the the newer generation, it's still relatively 'new' and the uses of it is still increasing as time progresses.

By 2023, 67% of the people in the world started using the internet (source: ITU).



02

DATA

Datasets & Graphs

Datasets

- The datasets had similar timeframes and formatting for column types
- The datasets originated from Gapminder.org, sources by ilo.org - Aged 15+ employment rate (%), and gapm.io- number of people not using internet.
- Both datasets can be freely accessed for everyone , no restrictions
- The datasets are about employment rates of people aged 15 and above, and people who do not use the internet, both are represented as countries and the period is in years

Datasets for Non-Internet Users

country <chr>	1990 <dbl>	1991 <dbl>	1992 <dbl>	1993 <dbl>	1994 <dbl>	1995 <dbl>	1996 <dbl>	1997 <dbl>	1998 <dbl>
Afghanistan	100	100	100	100	100	100.0	*NA*	*NA*	*NA*
Angola	100	100	100	100	100	100.0	100.0	100.0	100.0
Albania	100	100	100	100	100	100.0	100.0	100.0	99.9
Andorra	100	100	100	100	100	100.0	98.5	97.0	93.1
UAE	100	100	100	100	100	99.9	99.6	96.7	93.1
Argentina	100	100	100	100	100	99.9	99.9	99.7	99.2

Datasets for Employment Rate

country <chr>	1991 <dbl>	1992 <dbl>	1993 <dbl>	1994 <dbl>	1995 <dbl>	1996 <dbl>	1997 <dbl>	1998 <dbl>	1999 <dbl>
Afghanistan	42.5	42.5	42.5	42.5	42.4	42.4	42.3	42.2	42.2
Angola	75.0	75.0	75.2	75.1	74.9	74.9	74.8	74.7	74.6
Albania	57.8	58.2	56.8	55.7	54.1	53.3	54.5	53.8	52.7
UAE	71.8	72.2	72.9	73.4	73.8	73.3	73.1	73.3	73.7
Argentina	57.3	56.9	54.9	54.0	49.5	50.7	52.5	54.1	53.1
Armenia	56.0	55.5	53.1	52.0	51.4	49.6	48.4	48.8	47.5

Variable

For the 2 datasets, we grouped them into first-world countries and third-world countries!

- The x-axis represent years.
- The y-axis representing percentage.
- The left graph is for people that don’t use internet
- The right graph is for employment rates.

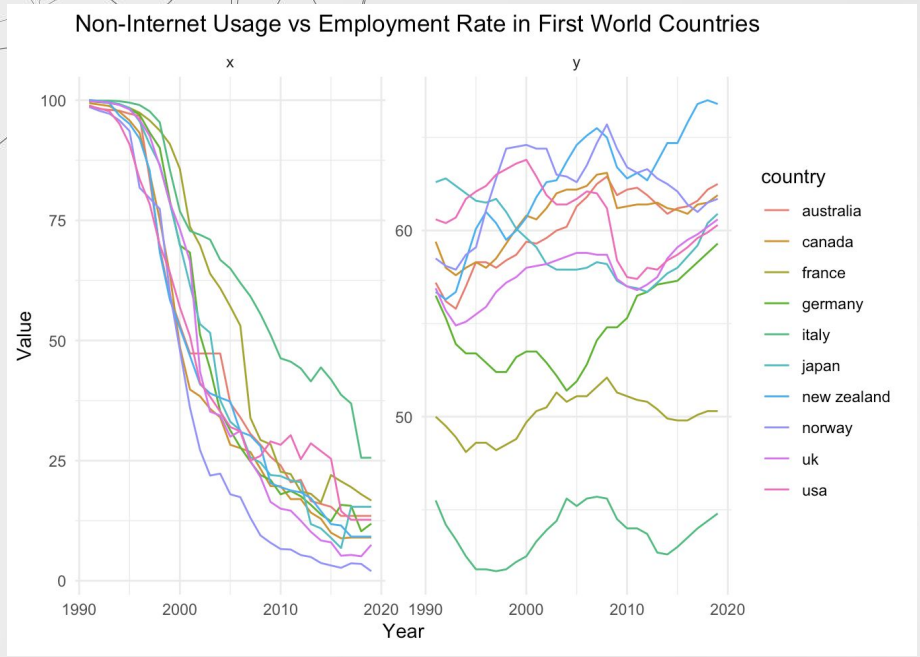
This is to show trends dependant on environment, and if they have relation.

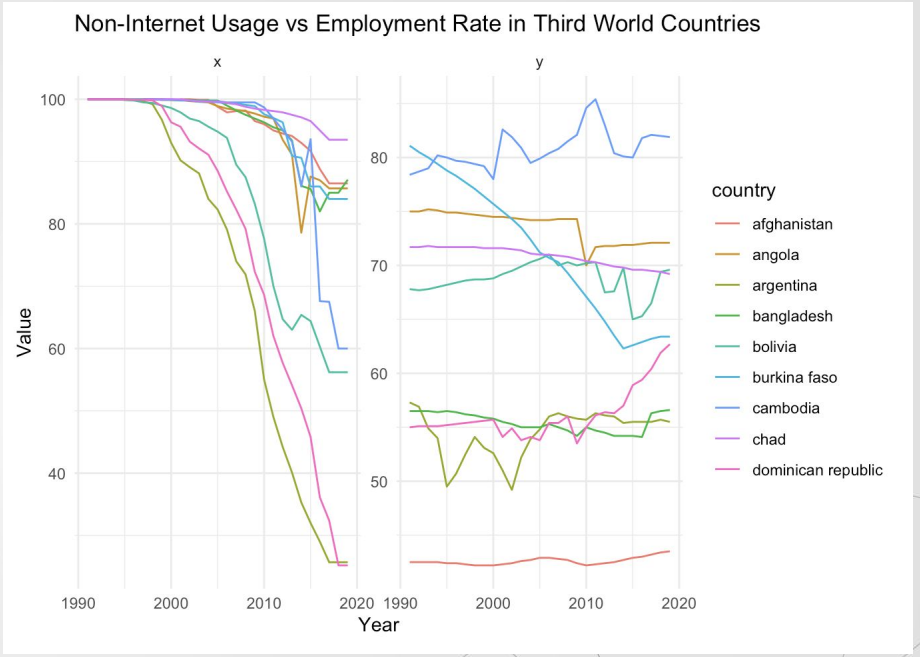
Combined Dataset

	country <chr>	1991.x <dbl>	1992.x <dbl>	1993.x <dbl>	1994.x <dbl>	1995.x <dbl>	1996.x <dbl>	1997.x <dbl>	1998.x <dbl>
1	afghanistan	100	100	100	100	100.0	NA	NA	NA
2	albania	100	100	100	100	100.0	100.0	100.0	99.9
3	algeria	100	100	100	100	100.0	100.0	100.0	100.0
4	angola	100	100	100	100	100.0	100.0	100.0	100.0
5	argentina	100	100	100	100	99.9	99.9	99.7	99.2
6	armenia	100	100	100	100	100.0	99.9	99.9	99.9

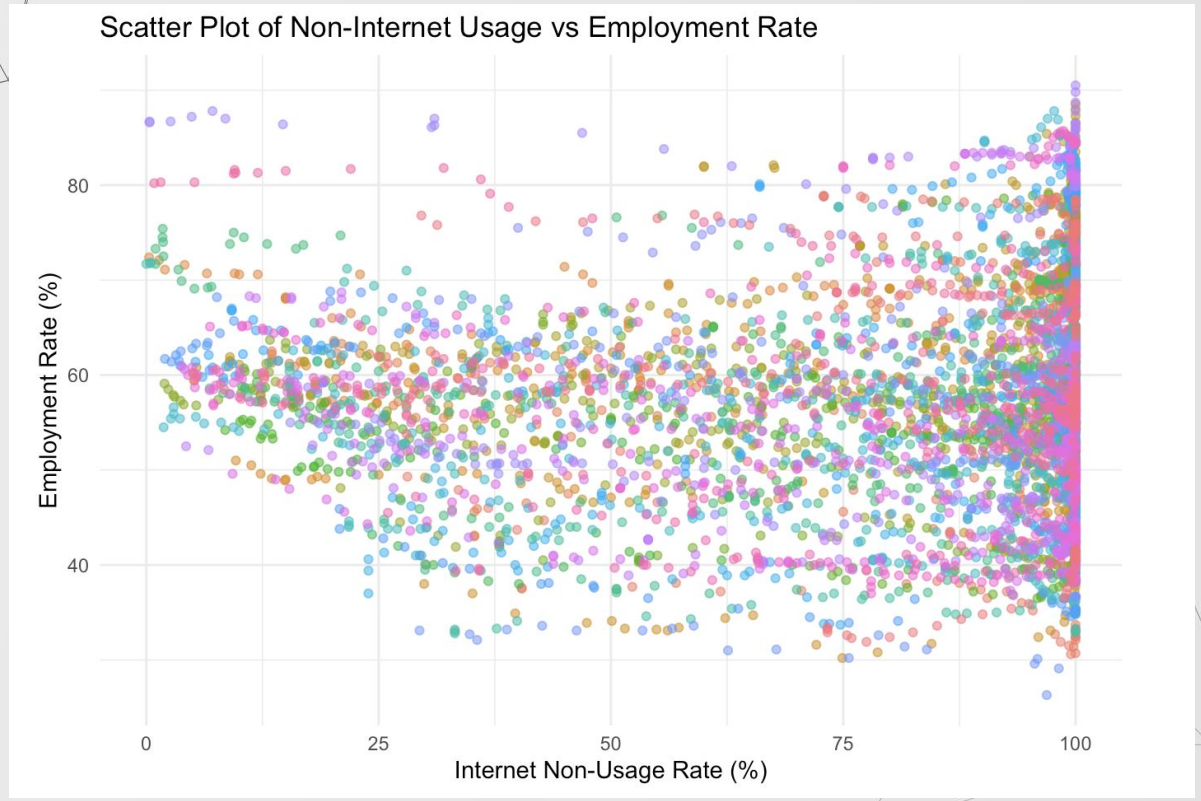
2000.y <dbl>	2001.y <dbl>	2002.y <dbl>	2003.y <dbl>	2004.y <dbl>	2005.y <dbl>	2006.y <dbl>	2007.y <dbl>	2008.y <dbl>	2009.y <dbl>
42.2	42.3	42.4	42.6	42.7	42.9	42.9	42.8	42.7	42.4
52.1	51.2	50.5	49.6	48.8	48.0	47.4	46.9	47.9	47.1
30.6	31.5	31.9	32.7	35.2	36.0	37.2	36.5	37.4	37.8
74.5	74.5	74.4	74.3	74.2	74.2	74.2	74.3	74.3	74.3
52.6	51.0	49.2	52.2	53.9	54.8	56.0	56.3	56.0	55.8
47.1	47.0	46.7	46.5	46.5	46.6	46.9	47.4	45.5	43.8

Line Graphs


Non-Internet Usage vs Employment Rate in First World Countries
x
y
Value
Year
country
australia
canada
france
germany
italy
japan
new zealand
norway
uk
usa


Non-Internet Usage vs Employment Rate in Third World Countries
x
y
Value
Year
country
afghanistan
angola
argentina
bangladesh
bolivia
burkina faso
cambodia
chad
dominican republic

Scatter Plot



03

PROJECT METHODS

Manipulating Data
&
Graphing

Load data

```{r, message=FALSE}
library(tidyverse)
library(readr)
library(ggplot2)
library(dplyr)
```

Trim both datasets so time matches 1991 to 2019

```{r}
internet_df <- internet_df %>%
  select(country, '1991':'2019')
head(internet_df)

employment_df <- employment_df %>%
  select(country, '1991':'2019')
head(employment_df)
```

Data Cleaning

```{r}
internet_df[is.na(internet_df)] <- NA
head(internet_df)

employment_df[is.na(employment_df)] <- NA
head(employment_df)
```

Filter countries for both datasets

```{r}
internet_df$country <- tolower(internet_df$country)
employment_df$country <- tolower(employment_df$country)

common_countries <- intersect(internet_df$country, employment_df$country)

internet_filtered <- internet_df %>% filter(country %in% common_countries)
dim(internet_filtered)
employment_filtered <- employment_df %>% filter(country %in% common_countries)
dim(employment_filtered)
```

```
[1] 181  30
[1] 181  30
```

Merge data

```{r}
combined_df <- merge(internet_filtered, employment_filtered, by = "country")
head(combined_df)
```

Reshape data

```{r, message=FALSE}
library(tidyr)
library(dplyr)
library(ggplot2)

# Reshape the data to long format
combined_long <- combined_df %>%
    pivot_longer(cols = -country, names_to = "year_indicator", values_to = "value") %>%
    separate(year_indicator, into = c("year", "indicator"), sep = "\\.")

# Convert year to numeric
combined_long$year <- as.numeric(combined_long$year)
```

Filter Countries

```{r}
# Define the lists of countries
first_world_countries <- tolower(c("Australia", "Canada", "France", "Germany", "Italy",
                           "Japan", "New Zealand", "Norway", "UK", "USA"))
third_world_countries <- tolower(c("Afghanistan", "Angola", "Argentina", "Bangladesh", "Bolivia",
                           "Burkina Faso", "Cambodia", "Chad", "Congo", "Dominican Republic"))

# Filter the data
first_world_data <- combined_long %>%
  filter(country %in% first_world_countries)
first_world_data
third_world_data <- combined_long %>%
  filter(country %in% third_world_countries)
third_world_data
```

Process, sample, and technical

Process

The way we processed the data were first to:

1. Trim the columns so they matched (same years)
2. Remove anything that were ‘blank’ (NA)
3. Cut rows out that weren’t common to both datasets (common countries only), both rows ended up having the same rows and columns, and used the country column to merge it together.

Sample Selection

We thought about the development of the countries, their access to internet and a statistic that has always been relevant is employment rates! Also helps answer our research questions.

Decisions and library choices

We decided to use the packages we are familiar with, such as Readr, dplyr, ggplot2, which are all in tidyverse

04

RESULTS

Trends & Findings

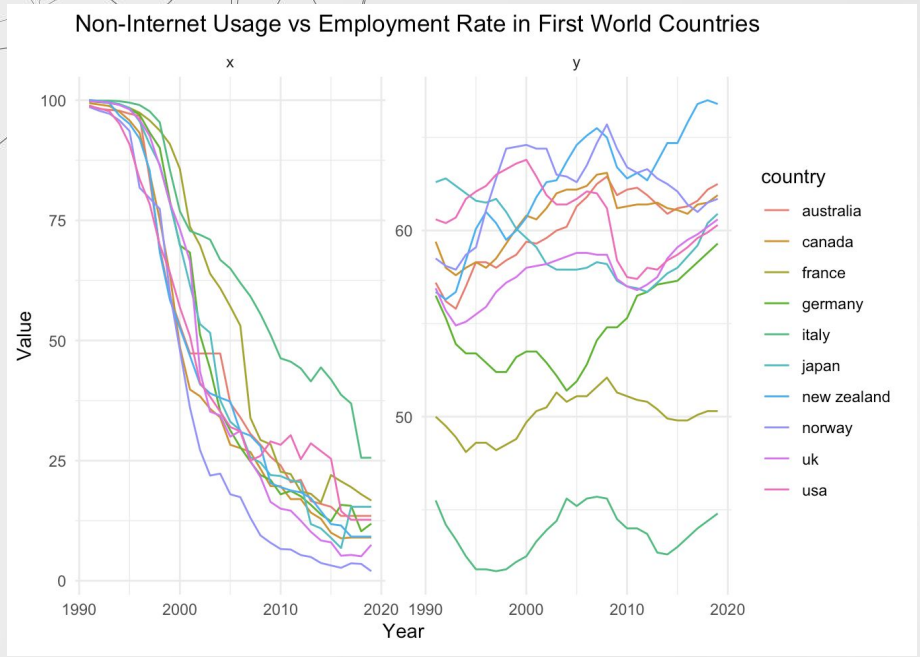
Findings

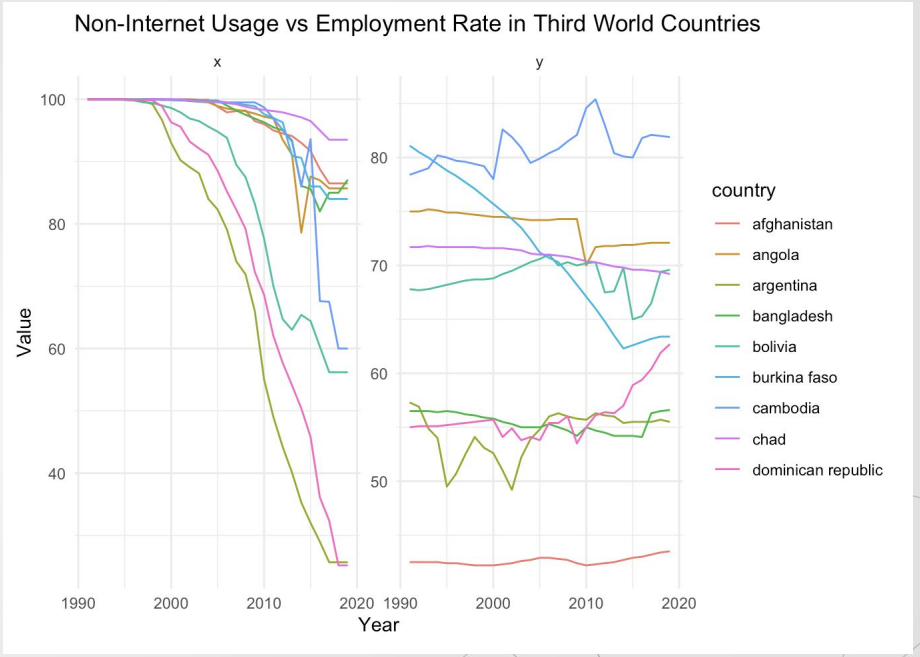
- There seems to be a correlation as more people gain access to the internet in first-world countries, there is also a jump in employment rates.

- In third-world countries, there seems to be little to no correlation in the chosen countries between internet access and employment rates.

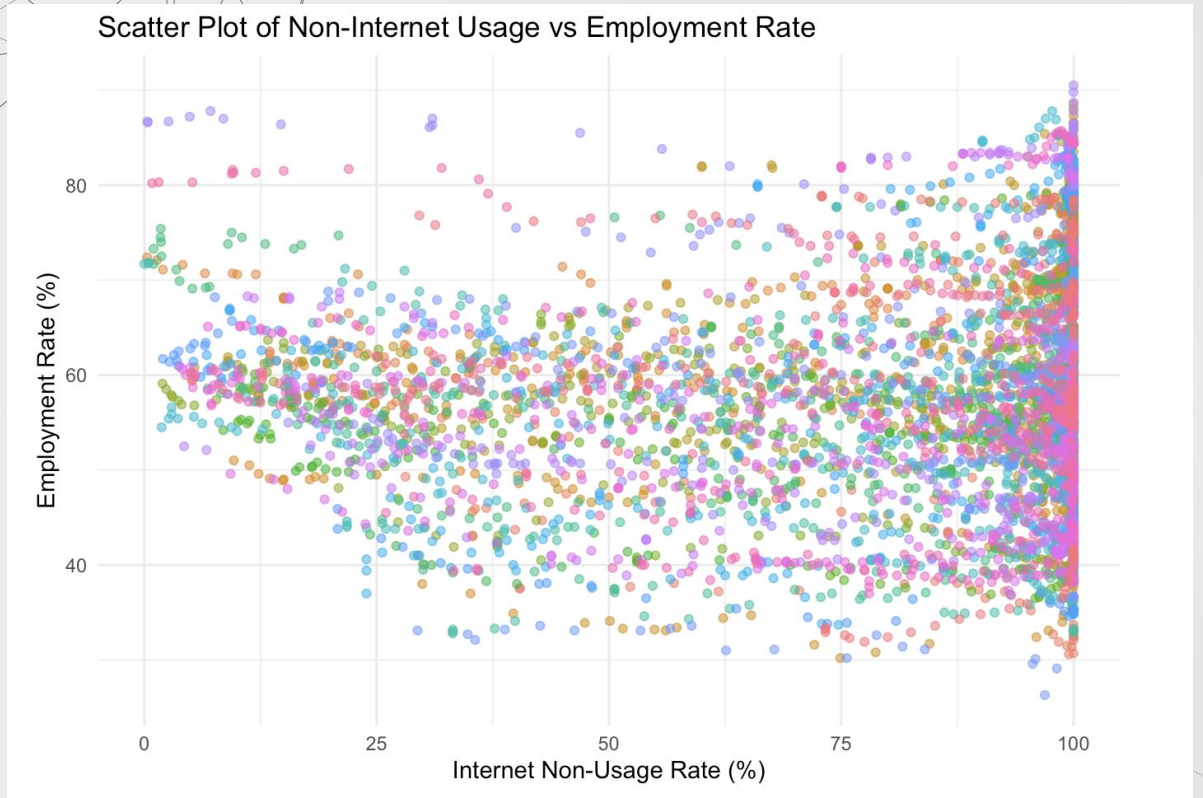
- Overall, when placed on the scatterplot, there doesn't seem to be any trends/correlation between internet access and employment rates overall.

Visualization of combined dataset





Overall Trend



- The data to the right shows no pattern, neither negative or positive correlations in any way.
- This is likely due to the majority of the world is still a developing country, and that some developed countries still have stigmas against remote work.

DISCUSSION

Interpretations & Limitations & Directions

05

Discuss

These findings help answer whether internet access has helped in employment rates prior to 2020 (which it has not overall, but has in selected developed countries)

The limitations to this is that we have no data 2020 and present day, limiting our ability to answer if COVID-19 has increased employment rates with increased internet usage.

In the future, we would need datasets that include the time period of 2019 to present day (COVID-19 period) ,which would give a more accurate representation of the impact that COVID-19 had to to increasing remote work .

SUMMARY

Our take away

06

Summary

The data presented from Gapminder had allowed us to see any current (up to 2019) trends in terms of internet non-usage, and employment rates in most countries. We were able to answer Lynnie's and Sean's questions! However, Alex's question wasn't answered due to lack of data

- Seemingly, internet usage does not correlate with employment rates in third-world/underdeveloped countries (from our selection pool)
- Internet usage increase in first-world/developed countries has correlation to employment rate, potentially meaning remote work is more accepted
- While there isn't information found in the data, further research found that remote work due to the COVID-19 pandemic is definitely correlated

Thank You!

(Question & Answer…)