# How Much of the World Has Access to the Internet ?

## This report presents the state of internet accessibility across the world by answering these specific questions:

1. What are the top 5 countries with the highest internet use in 2019?
2. How many people had internet access in those countries in 2019?
3. What are the top 5 countries with the highest internet use for each of the following regions: Middle East & North Africa, Latin America & Caribbean, East Asia & Pacific, Europe & Central Asia?
4. Creating a visualization for those four regions.
5. What is the correlation between internet usage (population share) and broadband subscriptions in 2019?
6. Summarize my findings

# The data

### internet

- "Entity" - The name of the country, region, or group.
- "Code" - Unique id for the country (null for other entities).
- "Year" - Year from 1990 to 2019.
- "Internet_usage" - The share of the entity's population who have used the internet in the last three months.

### people

- "Entity" - The name of the country, region, or group.
- "Code" - Unique id for the country (null for other entities).
- "Year" - Year from 1990 to 2020.
- "Users" - The number of people who have used the internet in the last three months for that country, region, or group.

### broadband

- "Entity" - The name of the country, region, or group.
- "Code" - Unique id for the country (null for other entities).
- "Year" - Year from 1998 to 2020.
- "Broadband_Subscriptions" - The number of fixed subscriptions to high-speed internet at downstream speeds >= 256 kbit/s for that country, region, or group.

```
library(readr)
library(dplyr)
library(DescTools)
library(patchwork)
library(tidyr)
library(ggplot2)
library(scales)
library(stats)


internet <- read_csv("data/internet.csv")%>%
            distinct()%>%
            drop_na()

people  <- read_csv("data/people.csv")%>%
           distinct()%>%
           drop_na()


broadband <- read_csv("data/broadband.csv")%>%
             distinct()%>%
             drop_na()
```

### 1. What are the top 5 countries with the highest internet use in 2019?

```
countries_with_highest_use <- internet%>%
                              group_by( Entity )%>%
                              filter( Year == 2019 )%>%
                              arrange( desc( Internet_Usage ) )%>%
                              mutate( Internet_Usage = paste( round( Internet_Usage,
                                      2 ),"%" ), Year = Year )%>%
                              head(5)%>%
                              select(-Code,-Year)

countries_with_highest_use
```

```
## # A tibble: 5 x 2
## # Groups:   Entity [5]
##   Entity                Internet_Usage
##   <chr>                 <chr>
## 1 Bahrain               99.7 %
## 2 Qatar                 99.65 %
## 3 Kuwait                99.54 %
## 4 United Arab Emirates 99.15 %
## 5 Denmark               98.05 %
```

### 2. How many people had internet access in those countries in 2019?

```
   num_of_people <- people%>%
                    filter( Year == 2019, Entity %in%
                            countries_with_highest_use$Entity )%>%
                    arrange( desc( Users ) )%>%
                    select(-Code,-Year)

   num_of_people
```

```
## # A tibble: 5 x 2
##   Entity                 Users
##   <chr>                  <dbl>
## 1 United Arab Emirates 9133361
## 2 Denmark              5682653
## 3 Kuwait               4420795
## 4 Qatar                2797495
## 5 Bahrain              1489735
```

```
   inner_join( countries_with_highest_use ,num_of_people , by=c( "Entity" ) )%>%
   arrange( desc( Users ) )
```

```
## # A tibble: 5 x 3
## # Groups:   Entity [5]
##   Entity                Internet_Usage   Users
##   <chr>                 <chr>            <dbl>
```

```
## 1 United Arab Emirates 99.15 %        9133361
## 2 Denmark              98.05 %        5682653
## 3 Kuwait               99.54 %        4420795
## 4 Qatar                99.65 %        2797495
## 5 Bahrain              99.7 %         1489735
```

**3. What are the top 5 countries with the highest internet use for each of the following regions: Middle East & North Africa,Latin America & Caribbean, East Asia & Pacific, Europe & Central Asia?**

```
our_regions <- c( 'Middle East & North Africa', 'Latin America & Caribbean',
                  'East Asia & Pacific', 'Europe & Central Asia' )
 regions <- d.countries%>%
            select( a3, region )%>%
            filter( region %in%  our_regions )

 internet_with_regions <- inner_join( internet, regions, by = c( "Code" = "a3" ) )



 #Middle East & North Africa
 top_Africa <- internet_with_regions%>%
               filter( region == "Middle East & North Africa", Year == 2019 )%>%
               arrange( desc( Internet_Usage ) )%>%
               head(5)

 top_Africa
```

```
## # A tibble: 5 x 5
##   Entity               Code   Year Internet_Usage region
##   <chr>                <chr> <dbl>          <dbl> <chr>
## 1 Bahrain              BHR    2019           99.7 Middle East & North Africa
## 2 Qatar                QAT    2019           99.7 Middle East & North Africa
## 3 Kuwait               KWT    2019           99.5 Middle East & North Africa
## 4 United Arab Emirates ARE    2019           99.1 Middle East & North Africa
## 5 Saudi Arabia         SAU    2019           95.7 Middle East & North Africa
```

```
 #Latin America & Caribbean

 top_Latin_America <- internet_with_regions%>%
                      filter( region == 'Latin America & Caribbean', Year == 2019 )%>%
                      arrange( desc( Internet_Usage ) )%>%
                      head(5)

 top_Latin_America
```

```
## # A tibble: 5 x 5
##   Entity      Code   Year Internet_Usage region
##   <chr>       <chr> <dbl>          <dbl> <chr>
## 1 Costa Rica  CRI    2019           81.2 Latin America & Caribbean
## 2 Uruguay     URY    2019           76.9 Latin America & Caribbean
```

```
## 3 Puerto Rico PRI    2019           70.9 Latin America & Caribbean
## 4 Mexico      MEX    2019           70.1 Latin America & Caribbean
## 5 Paraguay    PRY    2019           68.5 Latin America & Caribbean
```

```
    #East Asia & Pacific
    top_East_Asia <- internet_with_regions%>%
                     filter( region == 'East Asia & Pacific', Year == 2019 )%>%
                     arrange( desc( Internet_Usage ) )%>%
                     head(5)

    top_East_Asia
```

```
## # A tibble: 5 x 5
##   Entity      Code   Year Internet_Usage region
##   <chr>       <chr> <dbl>          <dbl> <chr>
## 1 South Korea KOR    2019           96.2 East Asia & Pacific
## 2 Brunei      BRN    2019           95   East Asia & Pacific
## 3 Hong Kong   HKG    2019           91.7 East Asia & Pacific
## 4 Singapore   SGP    2019           88.9 East Asia & Pacific
## 5 Macao       MAC    2019           86.5 East Asia & Pacific
```

```
    #'Europe & Central Asia'

    top_Europe_Central_Asia <- internet_with_regions%>%
                               filter( region == 'Europe & Central Asia', Year == 2019 )%>%
                               arrange( desc( Internet_Usage ) )%>%
                               head(5)

    top_Europe_Central_Asia
```

```
## # A tibble: 5 x 5
##   Entity      Code   Year Internet_Usage region
##   <chr>       <chr> <dbl>          <dbl> <chr>
## 1 Denmark     DNK    2019           98.0 Europe & Central Asia
## 2 Norway      NOR    2019           98.0 Europe & Central Asia
## 3 Sweden      SWE    2019           94.5 Europe & Central Asia
## 4 Netherlands NLD    2019           93.3 Europe & Central Asia
## 5 Switzerland CHE    2019           93.1 Europe & Central Asia
```

## 4. Creating a visualization for those four regions

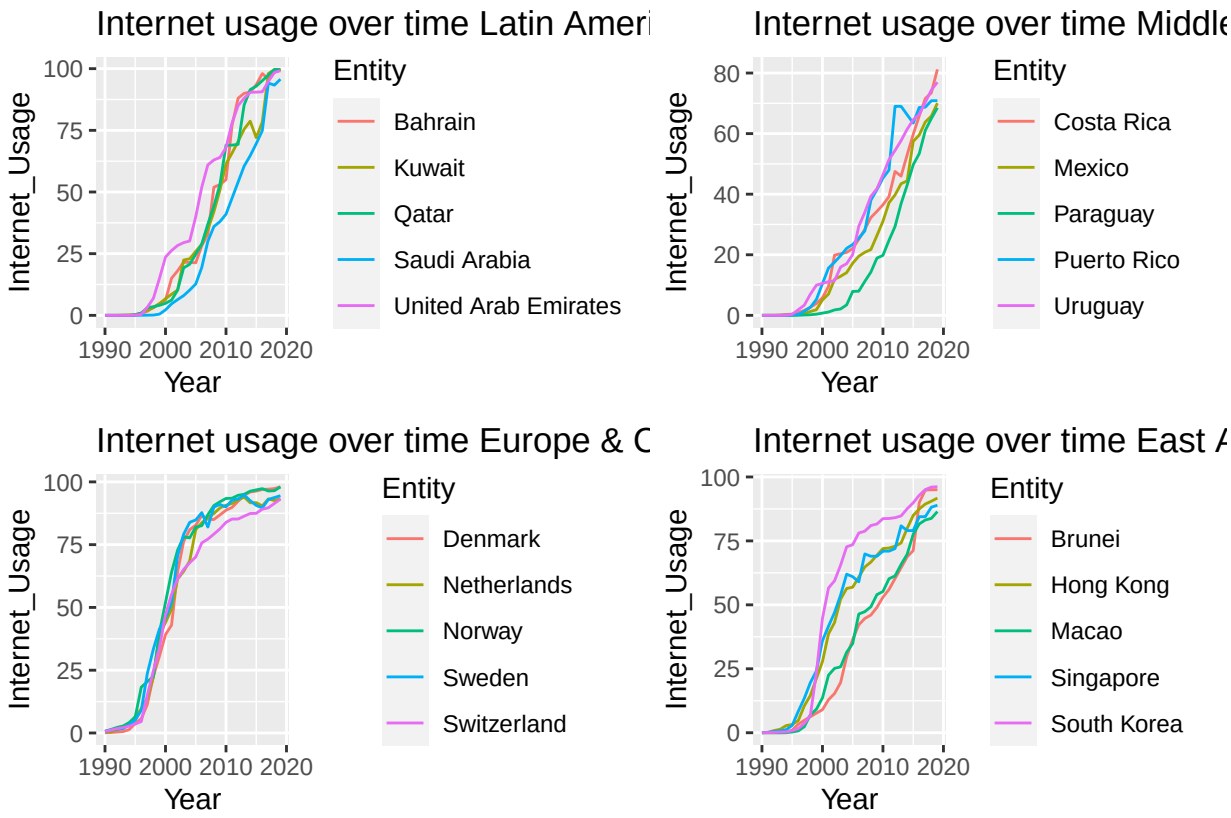
```
A <- internet_with_regions%>%
     filter( Entity %in% top_Africa$Entity )%>%
     ggplot( aes( Year, Internet_Usage, group=Entity, color=Entity ) )+geom_line()+
     ggtitle("Internet usage over time Latin America & Caribbean")


 B <- internet_with_regions%>%
      filter( Entity %in% top_Latin_America$Entity )%>%
      ggplot( aes( Year, Internet_Usage, group=Entity, color=Entity ) )+geom_line()+
      ggtitle( "Internet usage over time Middle East & North Africa" )
```

```
C <- internet_with_regions%>%
     filter( Entity %in% top_Europe_Central_Asia$Entity )%>%
     ggplot( aes( Year, Internet_Usage, group=Entity, color=Entity ) )+geom_line()+
     ggtitle( "Internet usage over time Europe & Central Asia" )


D <- internet_with_regions%>%
     filter( Entity %in% top_East_Asia$Entity )%>%
     ggplot( aes( Year, Internet_Usage, group=Entity, color=Entity ) )+geom_line()+
     ggtitle( "Internet usage over time East Asia & Pacific" )


  A+B+C+D
```



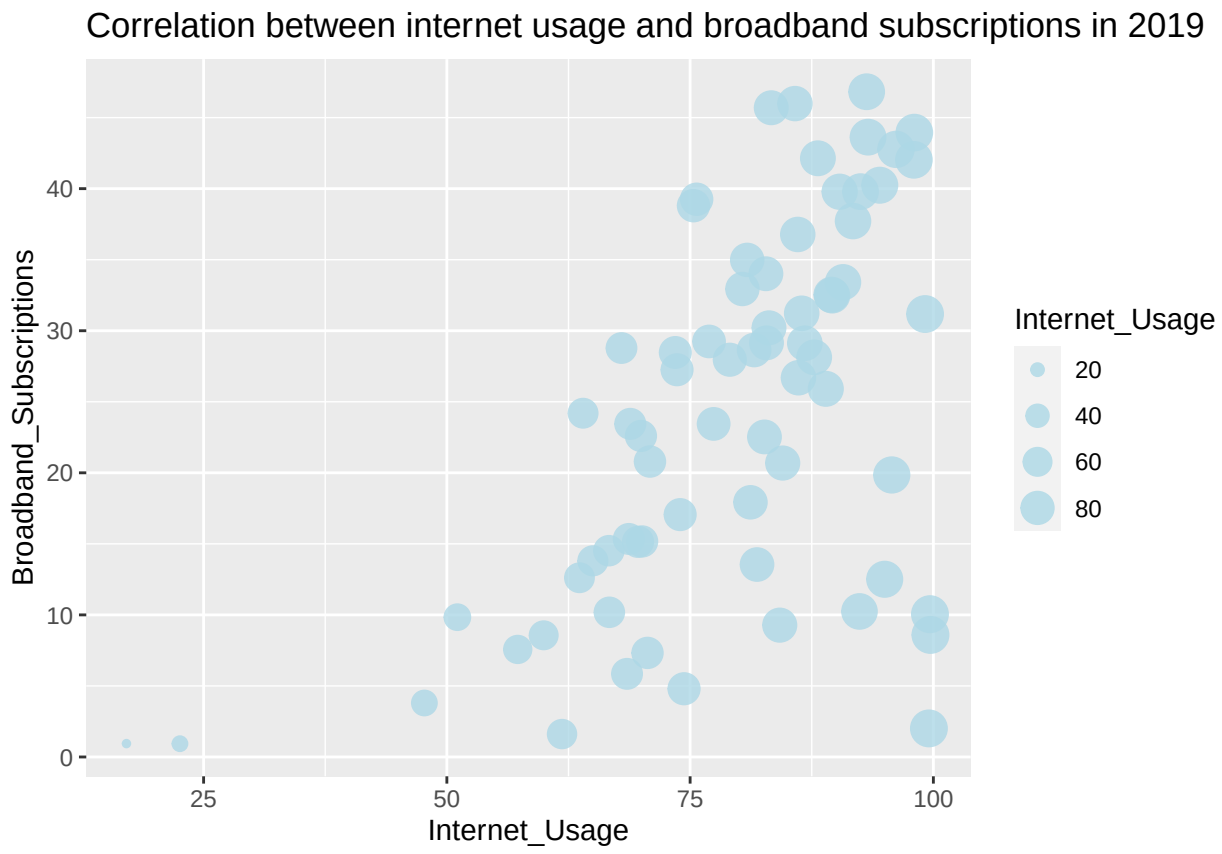

**5. What is the correlation between internet usage and broadband subscriptions in 2019?**

```
internet_with_broadband<-internet%>%
                         filter(Year == 2019)%>%
                         inner_join( broadband, by = c("Entity","Year"),
                         suffix = c( "_internet", "_broadband" ) )
```

```
    cor.test( internet_with_broadband$Internet_Usage,
        internet_with_broadband$Broadband_Subscriptions )
```

```
##
##  Pearson's product-moment correlation
##
## data:  internet_with_broadband$Internet_Usage and internet_with_broadband$Broadband_Subscriptions
## t = 5.4971, df = 67, p-value = 6.487e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3696086 0.7016177
## sample estimates:
##       cor
## 0.5575181
```

```
    ggplot(internet_with_broadband,aes( Internet_Usage, Broadband_Subscriptions,
    size = Internet_Usage ) )+geom_point( alpha = 0.8, color = "lightblue" )+
    ggtitle("Correlation between internet usage and broadband subscriptions in 2019")
```



## 6. Summarize my findings

**The top 5 countries with the highest internet use were:**

1. United Arab Emirates 99.15 percent of Internet use and 9133361 users
2. Denmark 98.05 percent of Internet use and 5682653 users
3. Kuwait 99.54 percent of Internet use and 4420795 users
4. Qatar 99.65 percent of Internet use and 2797495 users
5. Bahrain 99.7 percent of Internet use and 148973 users

**The top 5 countries with the highest internet use:**

- **Middle East & North Africa**: Bahrain, Qatar, Kuwait, United Arab Emirates, Saudi Arabia
- **Latin America & Caribbean**: Costa Rica, Uruguay, Puerto Rico, Mexico, Paraguay
- **East Asia & Pacific**: South Korea, Brunei, Hong Kong, Singapore, Macao
- **Europe & Central Asia**: Denmark, Norway, Sweden, Netherlands, Switzerlands

**The correlation between internet usage and broadband subscriptions in 2019**

Correlation coefficient between internet usage and broadband subscriptions is ***0.5575181***, which suggests a moderate positive correlation between the variables.

The ***t-value of 5.4971 with 67 degrees of freedom*** indicates a strong difference between the actual and theoretical correlation, and the very small ***p-value of 6.487e-07*** suggests that this difference is highly unlikely to be due to chance.

At ***0.05 level of significance, we reject the null hypothesis***, which states that the true correlation between them is equal to 0.