LSM Project

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(dplyr)
library(readxl)
library(repr)
```

```
## Warning: package 'repr' was built under R version 4.1.2
```

```
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 4.1.2

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

```
age_data<- read_excel("Project data.xlsx",
    sheet = "1", col_names = c("Age","2014","2015","2016","2017","2018","2019","2020","2021"), col_types
        "skip", "numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "numeric", "numeric"), skip = 7,
    n_max = 8)
```

```
age_data = age_data[-1,]

sex_data<- read_excel("Project data.xlsx",
    sheet = "1", col_names = c("Sex","2014","2015","2016","2017","2018","2019","2020","2021"), col_types
        "skip", "numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "numeric", "numeric"), skip = 15,
    n_max = 4)
sex_data = sex_data[-1,]

ethnic_data <- read_excel("Project data.xlsx",
    sheet = "4", col_names = c("Ethnicity","2014","2015","2016","2017","2018","2019","2020","2021"), co
        "skip", "numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "numeric", "numeric"), skip = 5)

eco_data <- read_excel("Project data.xlsx",
    sheet = "5", col_names = c("Activity","2014","2015","2016","2017","2018","2019","2020","2021"), col
        "skip", "numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "numeric", "numeric"), skip = 4,
    n_max = 9)
```

```
disability_data <- read_excel("Project data.xlsx",
    sheet = "3", col_names = c("AgeGroup","Disability","2017","2018","2019","2020","2021"), col_types =
        "text", "skip", "numeric", "numeric",
        "numeric", "numeric", "numeric"),
    skip = 6, n_max = 16)
```

```
age_sex <- read_excel("Project data.xlsx",
    sheet = "2", col_names = c("AgeGroup","Sex","2014","2015","2016","2017","2018","2019","2020","2021"
```

```
age_groups <- as.vector(age_data[,1])
sex<- as.vector(sex_data[,1])
#Data_byage <- Data_byage %>% remove_rownames %>% column_to_rownames(var="Age")
#Data_bysex <- Data_bysex %>% remove_rownames %>% column_to_rownames(var="Sex")
#by_ethnicity <- by_ethnicity %>% remove_rownames %>% column_to_rownames(var="Ethnicity")
#eco_activity <- eco_activity %>% remove_rownames %>% column_to_rownames(var="Activity")
```

1. Age Data

```
age_data %>% summary(Age)
```

```
##      Age                2014           2015           2016           2017
##  Length:7           Min.   : 898   Min.   :1057   Min.   :1371   Min.   :1534
##  Class :character   1st Qu.:4116   1st Qu.:4388   1st Qu.:4692   1st Qu.:5000
##  Mode  :character   Median :7145   Median :7088   Median :7075   Median :7074
##                     Mean   :5669   Mean   :5823   Mean   :6035   Mean   :6208
##                     3rd Qu.:7666   3rd Qu.:7840   3rd Qu.:7978   3rd Qu.:8095
##                     Max.   :8074   Max.   :8162   Max.   :8457   Max.   :8660
##       2018           2019           2020           2021
```

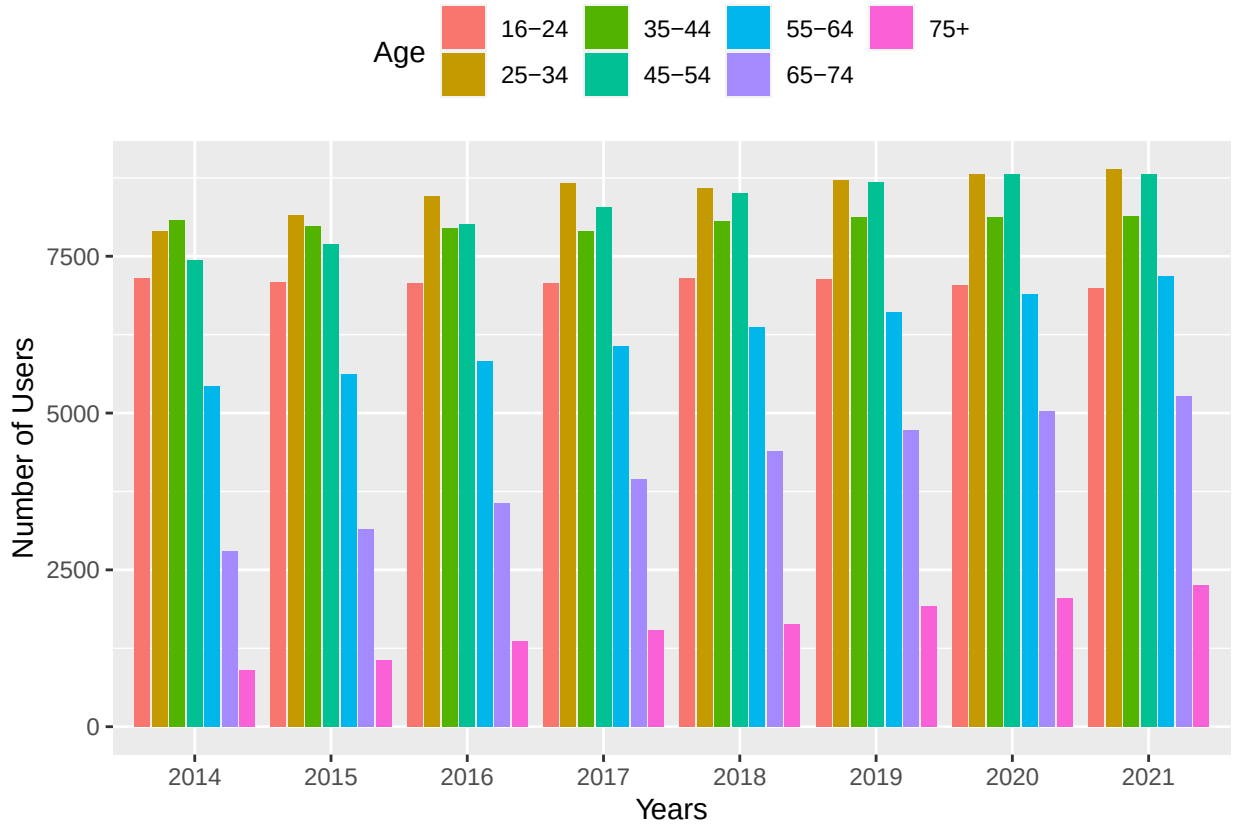
```
##  Min.   :1632   Min.   :1925   Min.   :2050   Min.   :2262
##  1st Qu.:5376   1st Qu.:5664   1st Qu.:5960   1st Qu.:6128
##  Median :7155   Median :7129   Median :7036   Median :7189
##  Mean   :6382   Mean   :6560   Mean   :6677   Mean   :6794
##  3rd Qu.:8276   3rd Qu.:8408   3rd Qu.:8460   3rd Qu.:8480
##  Max.   :8582   Max.   :8720   Max.   :8815   Max.   :8894
```

```
age_pivot <- age_data %>% pivot_longer(-Age,names_to = "Years", values_to = "Count")
age_pivot %>% summary()
```

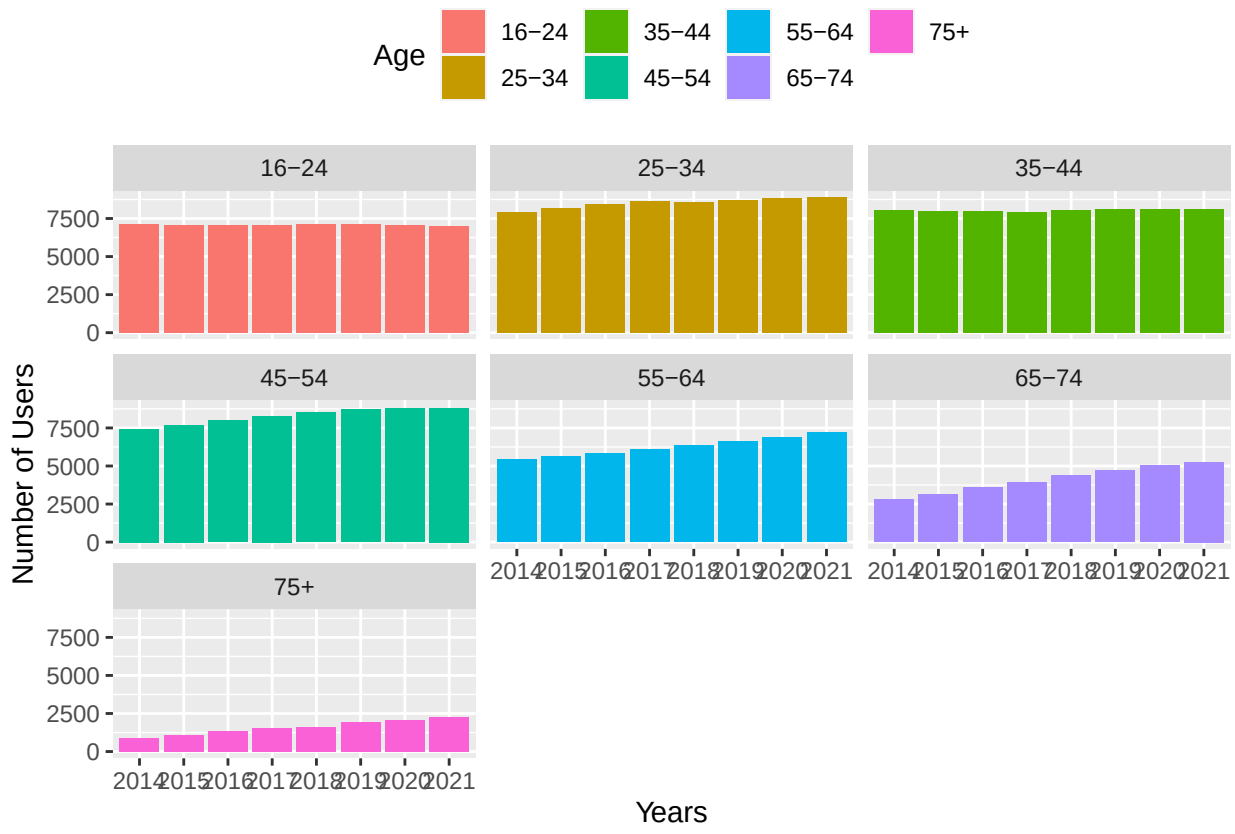
```
##      Age                Years               Count
##  Length:56          Length:56          Min.   : 898
##  Class :character   Class :character   1st Qu.:4954
##  Mode  :character   Mode  :character   Median :7108
##                                        Mean   :6268
##                                        3rd Qu.:8121
##                                        Max.   :8894
```

Visualizing the age_data

```
p <- ggplot(age_pivot,aes(x=Years,y = Count,fill= Age)) + geom_col(position = "dodge2") + labs(x="Years
  theme(legend.position = "top")
ggsave("age_data.png", p, width = 15, height = 10)
p
```

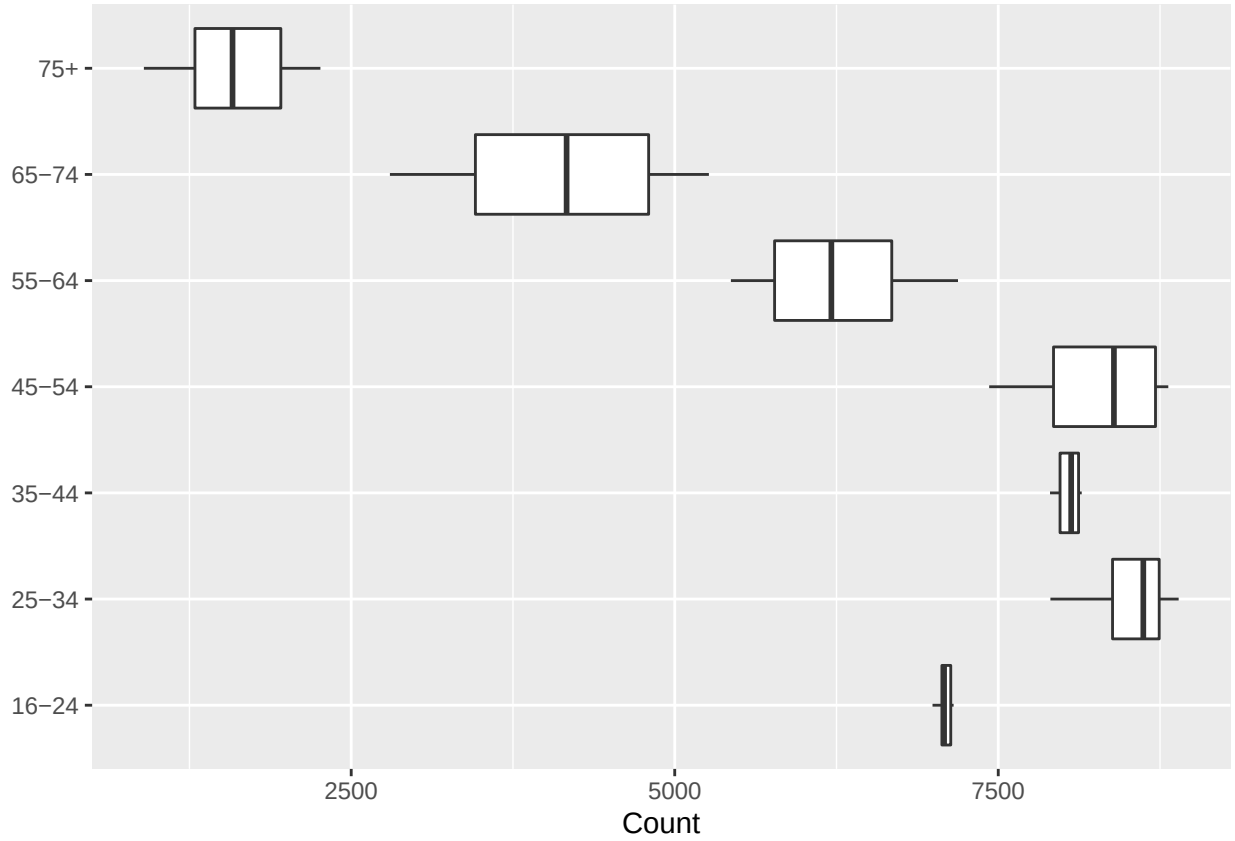


```
p <- ggplot(age_pivot,aes(x=Years,y = Count,fill= Age)) + geom_col(position = "dodge2") + labs(x="Years
  theme(legend.position = "top")
ggsave("age_data_grouped.png", p, width = 15, height = 10)
p
```



Boxplot for same data

```
p1 <- ggplot(age_pivot,aes(x = Age, y = Count)) + geom_boxplot() + labs(x=NULL) + coord_flip()+theme(leg
ggsave("age_boxplot.png", p1, width = 15, height = 10)
p1
```




```
age_data %>% select(`2014`,`2015`,`2016`,`2017`,`2018`,`2019`,`2020`,`2021`) %>% rowSums()
```

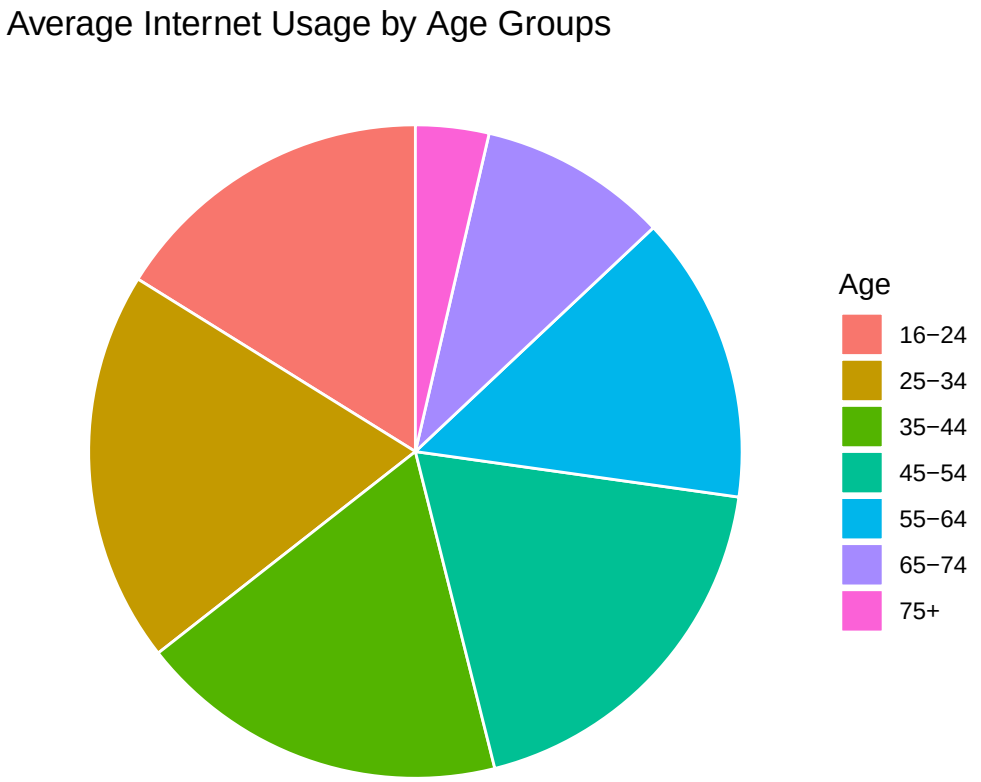
```
## [1] 56694 68193 64357 66220 49984 32859 12729
```

```
years <- c("2014","2015","2016","2017","2018","2019","2020","2021")
age_mean <- age_data %>% mutate(mean_users = rowMeans(.[, years]))

#df1 <- pivotted %>% group_by(Age) %>% mutate(mean_users = mean(Count))
age_mean <- age_mean %>% select(Age,mean_users)
```

Pie Chart for Mean Value

```
pie <- ggplot(age_mean, aes(x="", y=mean_users, fill=Age)) +
  geom_bar(stat="identity", width=1, color="white") + labs(title = "Average Internet Usage by Age Groups
  coord_polar("y", start=0) + theme_void()
ggsave("age_pie.png", pie, width = 15, height = 10)
pie
```



Anova test for Data by Age

```
#t.test(data = df1, mean_users ~ Age)
age_aov <- aov(Count ~ Age, data = age_pivot)
summary(age_aov)
```

```
##             Df    Sum Sq  Mean Sq F value Pr(>F)
## Age          6 315974373 52662396   203.4 <2e-16 ***
## Residuals   49  12687109   258921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Splitting Data by Age Groups

```
age_groups
```

```
## # A tibble: 7 x 1
##   Age
##   <chr>
## 1 16-24
## 2 25-34
## 3 35-44
## 4 45-54
## 5 55-64
## 6 65-74
## 7 75+
```

```r
age_data1 <- age_data %>% filter(Age == "16-24") %>% pivot_longer(-Age, names_to = "Years", values_to =
age_data2 <- age_data %>% filter(Age == "25-34") %>% pivot_longer(-Age, names_to = "Years", values_to =
age_data3 <- age_data %>% filter(Age == "35-44") %>% pivot_longer(-Age, names_to = "Years", values_to =
age_data4 <- age_data %>% filter(Age == "45-54") %>% pivot_longer(-Age, names_to = "Years", values_to =
age_data5 <- age_data %>% filter(Age == "55-64") %>% pivot_longer(-Age, names_to = "Years", values_to =
age_data6 <- age_data %>% filter(Age == "65-74") %>% pivot_longer(-Age, names_to = "Years", values_to =
age_data7 <- age_data %>% filter(Age == "75+") %>% pivot_longer(-Age, names_to = "Years", values_to = "

age_data1
```

```
## # A tibble: 8 x 3
##   Age   Years Users
##   <chr> <chr> <dbl>
## 1 16-24 2014   7145
## 2 16-24 2015   7088
## 3 16-24 2016   7075
## 4 16-24 2017   7074
## 5 16-24 2018   7155
## 6 16-24 2019   7129
## 7 16-24 2020   7036
## 8 16-24 2021   6992
```

```r
age_data2
```

```
## # A tibble: 8 x 3
##   Age   Years Users
##   <chr> <chr> <dbl>
## 1 25-34 2014   7903
## 2 25-34 2015   8162
## 3 25-34 2016   8457
## 4 25-34 2017   8660
## 5 25-34 2018   8582
## 6 25-34 2019   8720
## 7 25-34 2020   8815
## 8 25-34 2021   8894
```

Age Group 1: 16 -24 Applying Linear Regression

```r
age_data1_lm <- lm(Years ~ Users, data=age_data1)
summary(age_data1_lm)
```

```
##
## Call:
## lm(formula = Years ~ Users, data = age_data1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.469 -1.856  0.152  1.473  2.560
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 2195.22823  104.60898  20.985 7.63e-07 ***
## Users         -0.02508    0.01476  -1.699     0.14
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.174 on 6 degrees of freedom
## Multiple R-squared:  0.3248, Adjusted R-squared:  0.2123
## F-statistic: 2.887 on 1 and 6 DF,  p-value: 0.1402
```

```
#var(age_data1$)
```

Age Group 2: 25-34 Applying Linear Regression

```
age_data2_lm <- lm(Years ~ Users, data=age_data2)
summary(age_data2_lm)
```

```
##
## Call:
## lm(formula = Years ~ Users, data = age_data2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4232 -0.2905  0.1379  0.5728  0.9867
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.960e+03  8.623e+00 227.252  4.9e-13 ***
## Users       6.795e-03  1.011e-03   6.722 0.000527 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9059 on 6 degrees of freedom
## Multiple R-squared:  0.8828, Adjusted R-squared:  0.8632
## F-statistic: 45.18 on 1 and 6 DF,  p-value: 0.0005273
```

Age Group 3: 35-44 Applying Linear Regression

```
age_data3_lm <- lm(Years ~ Users, data=age_data3)
summary(age_data3_lm)
```

```
##
## Call:
## lm(formula = Years ~ Users, data = age_data3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9791 -0.3781  0.2437  1.4422  1.8631
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.886e+03  7.182e+01  26.265 2.01e-07 ***
## Users       1.631e-02  8.927e-03   1.827    0.117
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.121 on 6 degrees of freedom
## Multiple R-squared:  0.3574, Adjusted R-squared:  0.2503
## F-statistic: 3.337 on 1 and 6 DF,  p-value: 0.1175
```

Age Group 4: 45-54 Applying Linear Regression

```
age_data4_lm <- lm(Years ~ Users, data=age_data4)
summary(age_data4_lm)
```

```
##
## Call:
## lm(formula = Years ~ Users, data = age_data4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.55713 -0.40231 -0.07819  0.21867  1.04782
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.980e+03  3.499e+00  565.85 2.06e-15 ***
## Users       4.571e-03  4.219e-04   10.83 3.66e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5835 on 6 degrees of freedom
## Multiple R-squared:  0.9514, Adjusted R-squared:  0.9433
## F-statistic: 117.4 on 1 and 6 DF,  p-value: 3.664e-05
```

Age Group 5: 55-64 Applying Linear Regression

```
age_data5_lm <- lm(Years ~ Users, data=age_data5)
summary(age_data5_lm)
```

```
##
## Call:
## lm(formula = Years ~ Users, data = age_data5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30220 -0.08564  0.02092  0.11162  0.23856
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.993e+03  7.503e-01 2656.07  < 2e-16 ***
## Users       3.928e-03  1.196e-04   32.85 5.29e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1967 on 6 degrees of freedom
## Multiple R-squared:  0.9945, Adjusted R-squared:  0.9936
## F-statistic:  1079 on 1 and 6 DF,  p-value: 5.29e-08
```

Age Group 6: 65-74 Applying Linear Regression

```
age_data6_lm <- lm(Years ~ Users, data=age_data6)
summary(age_data6_lm)
```

```
##
## Call:
## lm(formula = Years ~ Users, data = age_data6)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.27208 -0.07410 -0.01665  0.08247  0.34032
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.006e+03  3.545e-01  5659.2  < 2e-16 ***
## Users       2.732e-03  8.458e-05    32.3 5.86e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2001 on 6 degrees of freedom
## Multiple R-squared:  0.9943, Adjusted R-squared:  0.9933
## F-statistic:  1043 on 1 and 6 DF,  p-value: 5.856e-08
```

Age Group 7: 75+ Applying Linear Regression

```
age_data7_lm <- lm(Years ~ Users, data=age_data7)
summary(age_data7_lm)
```

```
##
## Call:
## lm(formula = Years ~ Users, data = age_data7)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37441 -0.20740  0.05689  0.17299  0.29099
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.009e+03  3.361e-01 5979.28  < 2e-16 ***
## Users       5.113e-03  2.034e-04   25.14 2.61e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2565 on 6 degrees of freedom
## Multiple R-squared:  0.9906, Adjusted R-squared:  0.989
## F-statistic: 632.2 on 1 and 6 DF,  p-value: 2.606e-07
```

```
age_lm <- lm(formula = Count ~ Age, data = age_pivot)
summary(age_lm)
```

```
##
```

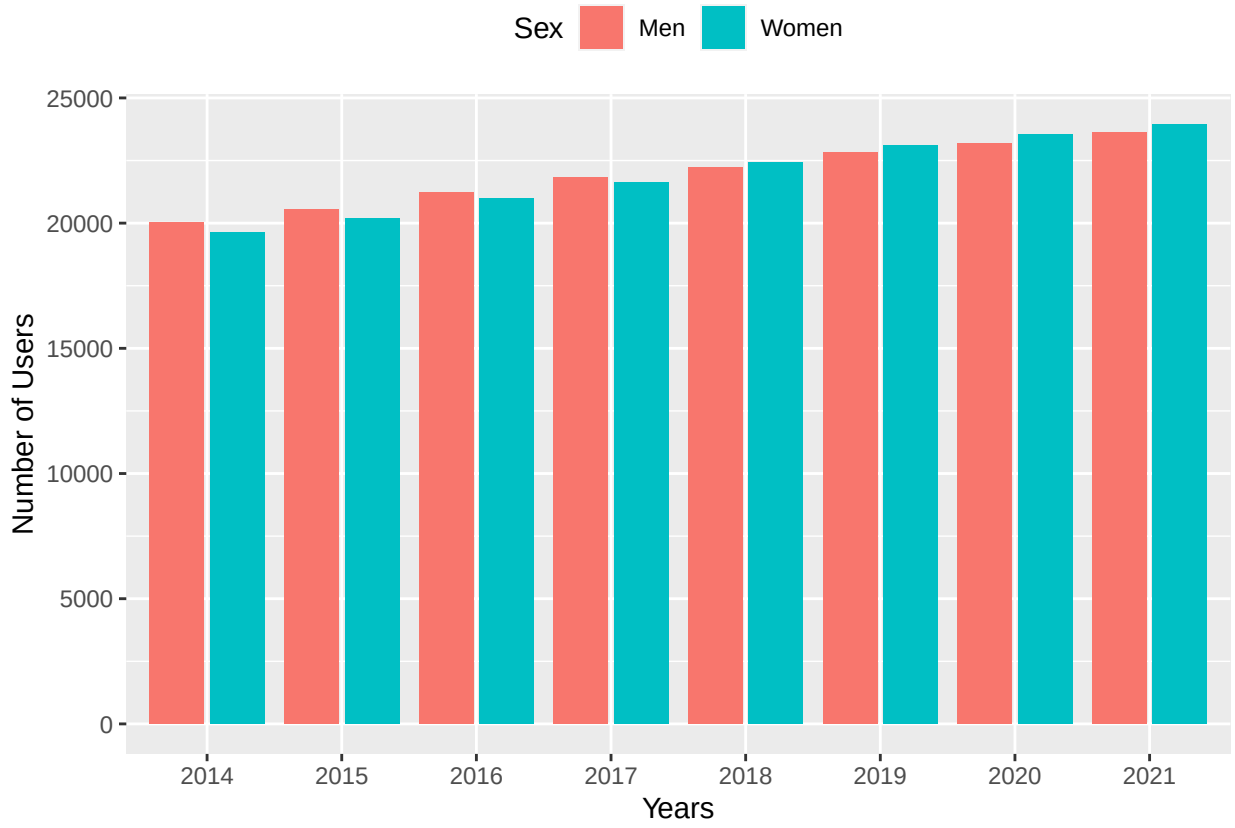
```
## Call:
## lm(formula = Count ~ Age, data = age_pivot)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1308.37  -196.03    20.94   284.69  1156.62
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7086.8      179.9  39.392  < 2e-16 ***
## Age25-34      1437.4      254.4   5.650 8.07e-07 ***
## Age35-44       957.9      254.4   3.765 0.000447 ***
## Age45-54      1190.7      254.4   4.680 2.29e-05 ***
## Age55-64      -838.7      254.4  -3.297 0.001825 **
## Age65-74     -2979.4      254.4 -11.710 8.26e-16 ***
## Age75+       -5495.6      254.4 -21.600  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 508.8 on 49 degrees of freedom
## Multiple R-squared:  0.9614, Adjusted R-squared:  0.9567
## F-statistic: 203.4 on 6 and 49 DF,  p-value: < 2.2e-16
```

2. Sex Data Data on the basis of Sex has exactly 2 levels, hence we can use t-test

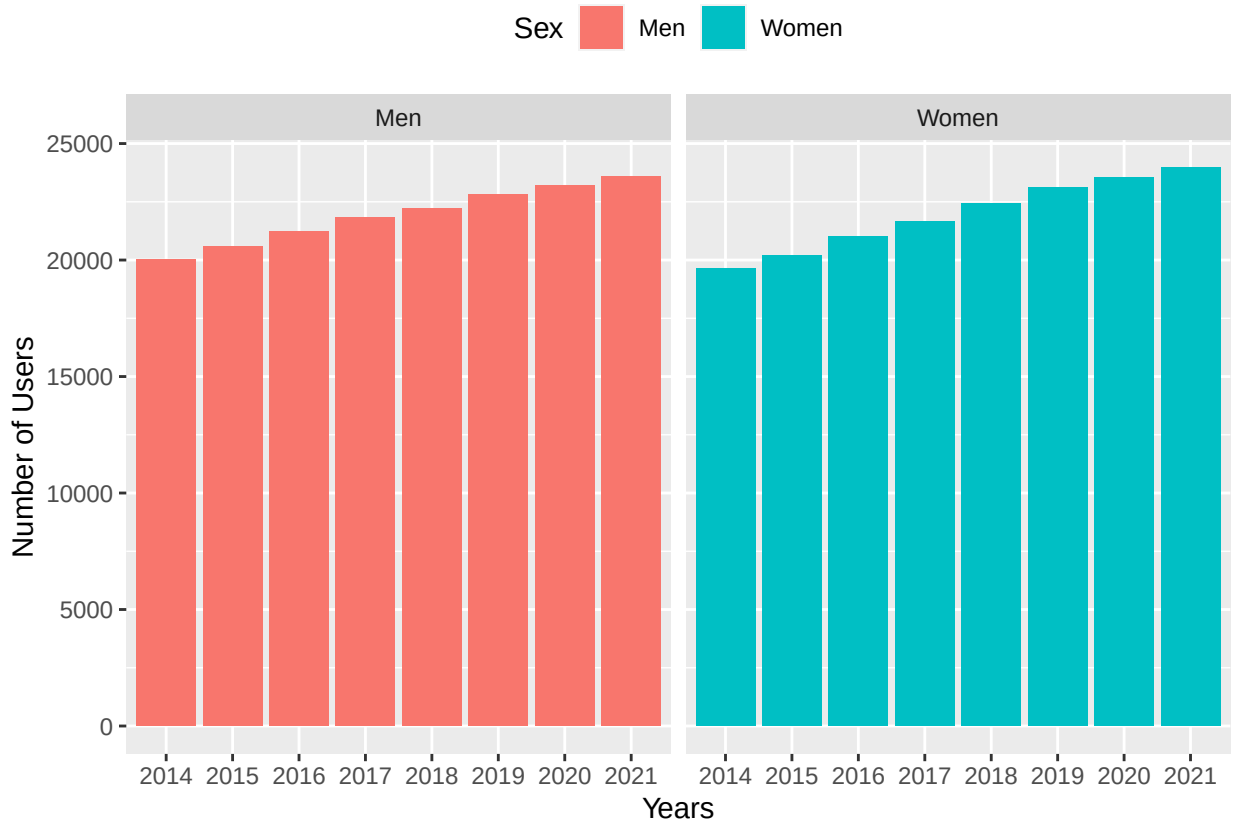
```
sex_pivot <- sex_data %>% pivot_longer(-Sex, names_to = "Years", values_to = "no_of_users")
sex_pivot
```

```
## # A tibble: 16 x 3
##    Sex   Years no_of_users
##    <chr> <chr>       <dbl>
##  1 Men   2014        20039
##  2 Men   2015        20564
##  3 Men   2016        21242
##  4 Men   2017        21814
##  5 Men   2018        22229
##  6 Men   2019        22812
##  7 Men   2020        23200
##  8 Men   2021        23606
##  9 Women 2014        19645
## 10 Women 2015        20200
## 11 Women 2016        21001
## 12 Women 2017        21643
## 13 Women 2018        22442
## 14 Women 2019        23105
## 15 Women 2020        23542
## 16 Women 2021        23954
```

```
p <- ggplot(sex_pivot,aes(x=Years, y = no_of_users,fill = Sex)) + geom_col(position = "dodge2") + labs(
  theme(legend.position = "top")
ggsave("sex_data.png", p, width = 15, height = 10)
p
```

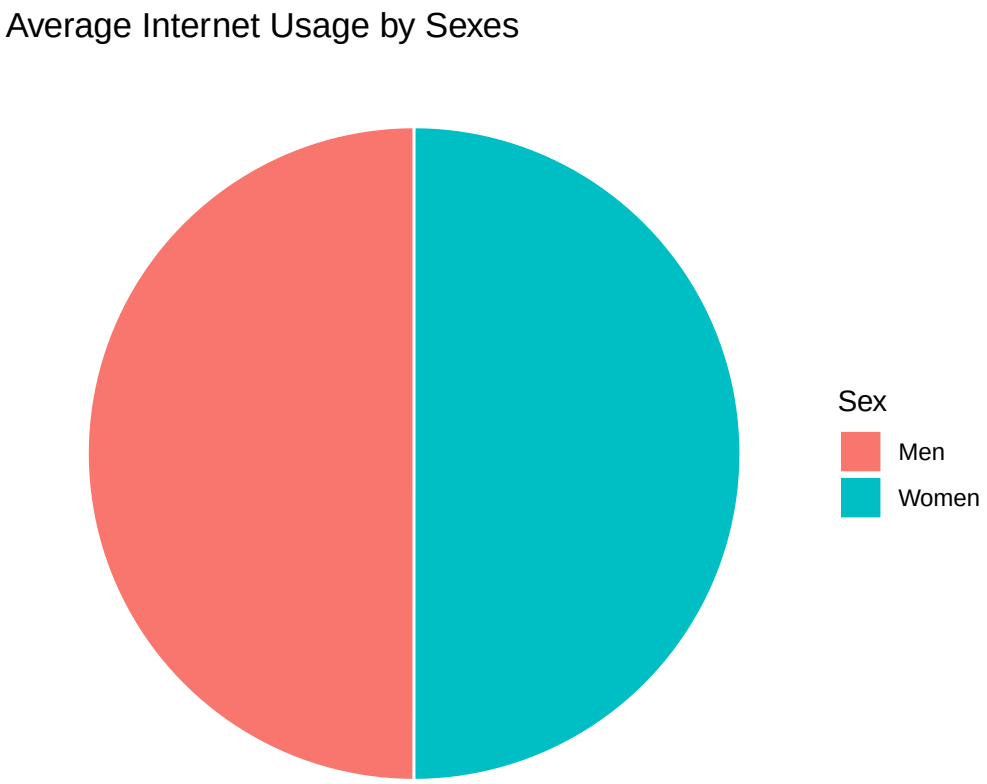



```
p <- ggplot(sex_pivot,aes(x=Years, y = no_of_users,fill = Sex)) + geom_col(position = "dodge2") + labs(
  theme(legend.position = "top")
ggsave("sex_data_grouped.png", p, width = 15, height = 10)
p
```




```
sex_groups <- c("Male","Female")
sex_mean <- sex_data %>% mutate(mean_users = rowMeans(.[, years]))
sex_mean <- sex_mean %>% select(Sex, mean_users)
#df2 <- df2 %>% select(Sex,m)
```

```
pie <- ggplot(sex_mean, aes(x="", y=mean_users, fill=Sex)) +
  geom_bar(stat="identity", width=1, color="white") + labs(title = "Average Internet Usage by Sexes") +
  coord_polar("y", start=0) + theme_void()
ggsave("sex_pie.png",pie, width = 15, height = 10)
pie
```




```
t.test(data = sex_pivot, no_of_users ~ Sex)
```

```
##
##  Welch Two Sample t-test
##
## data:  no_of_users by Sex
## t = -0.0045387, df = 13.361, p-value = 0.9964
## alternative hypothesis: true difference in means between group Men and group Women is not equal to 0
## 95 percent confidence interval:
##  -1545.968  1539.468
## sample estimates:
##   mean in group Men mean in group Women
##            21938.25            21941.50
```

```
sex_anova <- aov(no_of_users ~ Sex, data = sex_pivot)
summary(sex_anova)
```

```
##             Df   Sum Sq Mean Sq F value Pr(>F)
## Sex          1       42      42       0  0.996
## Residuals   14 28713659 2050976
```

3. Ethnicity Data

```
ethnic_pivot <- ethnic_data %>% pivot_longer(-Ethnicity,names_to = "Years", values_to = "Users")
ethnic_pivot
```

```
## # A tibble: 72 x 3
##    Ethnicity                          Years Users
##    <chr>                              <chr> <dbl>
##  1 White                              2014  35546
##  2 White                              2015  36430
##  3 White                              2016  37585
##  4 White                              2017  38601
##  5 White                              2018  39498
##  6 White                              2019  40526
##  7 White                              2020  40885
##  8 White                              2021  41825
##  9 Mixed/multiple ethnic background 2014    343
## 10 Mixed/multiple ethnic background 2015    401
## # ... with 62 more rows
```

Visualizing Ethnic Data

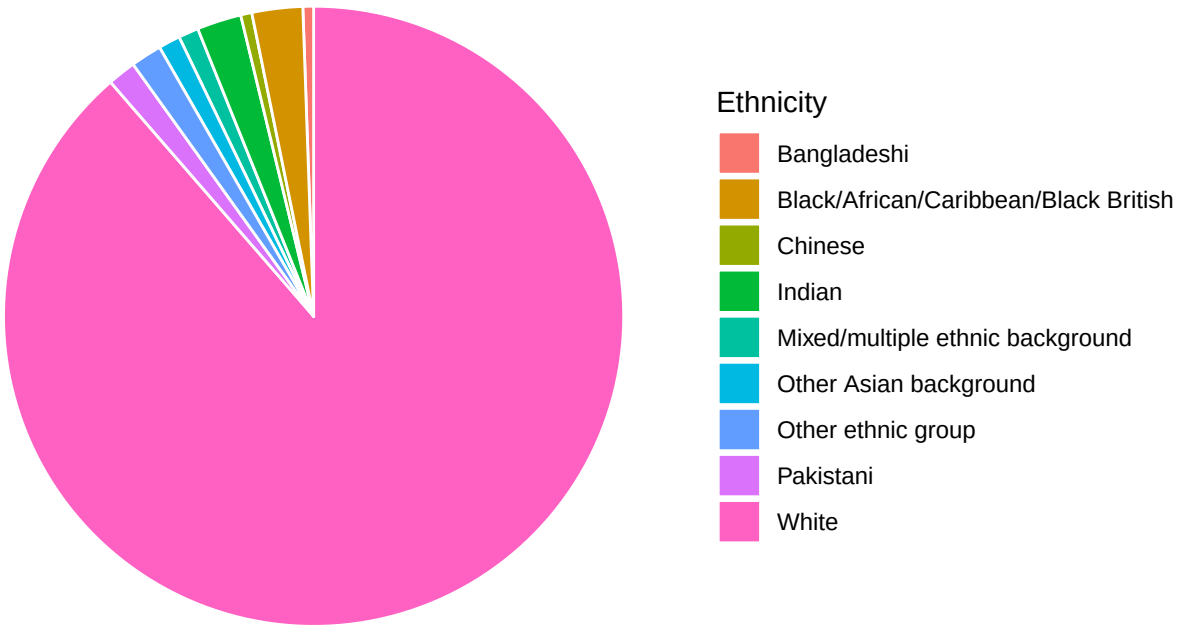
```
options(repr.plot.width = 20, repr.plot.height = 12)
e <- ggplot(ethnic_pivot,aes(x=Years, y = Users, fill =Ethnicity)) + geom_col(position = "dodge2") + lab
#options(repr.e.width=10,repr.e.height=8)
ggsave("ethnic_data.png", e, width = 15, height = 10)
e
```

Mean of Ethnic Data

```
ethnic_mean <- ethnic_data %>% mutate(mean_users = rowMeans(.[, years]))
ethnic_mean <- ethnic_mean %>% select(Ethnicity, mean_users)
```

Pie Chart

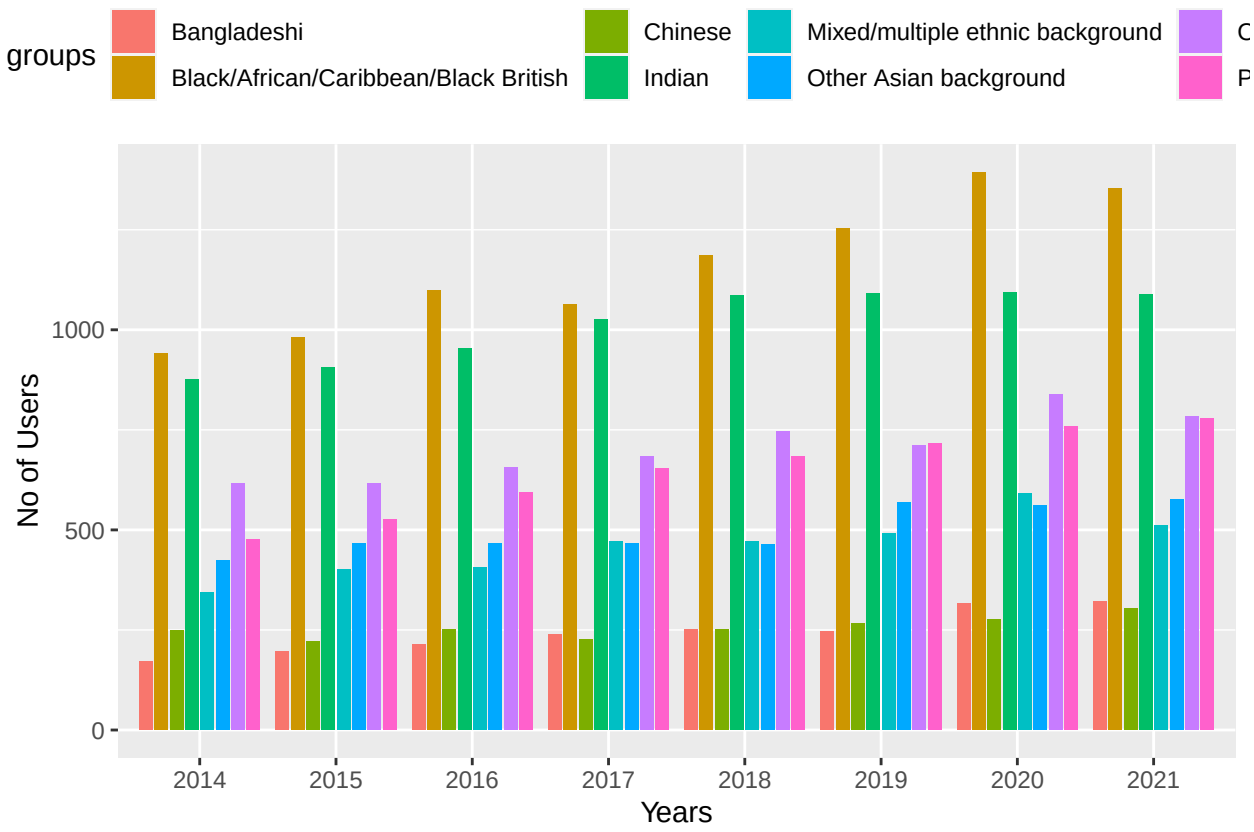
```
pie <- ggplot(ethnic_mean, aes(x="", y=mean_users, fill=Ethnicity)) +
  geom_bar(stat="identity", width=1, color="white") + labs(title = "Average Internet Usage by Ethnicity
  coord_polar("y", start=0) + theme_void()
ggsave("ethnic_pie.png",pie, width = 15, height = 10)
pie
```



Removing "White" Ethnic group

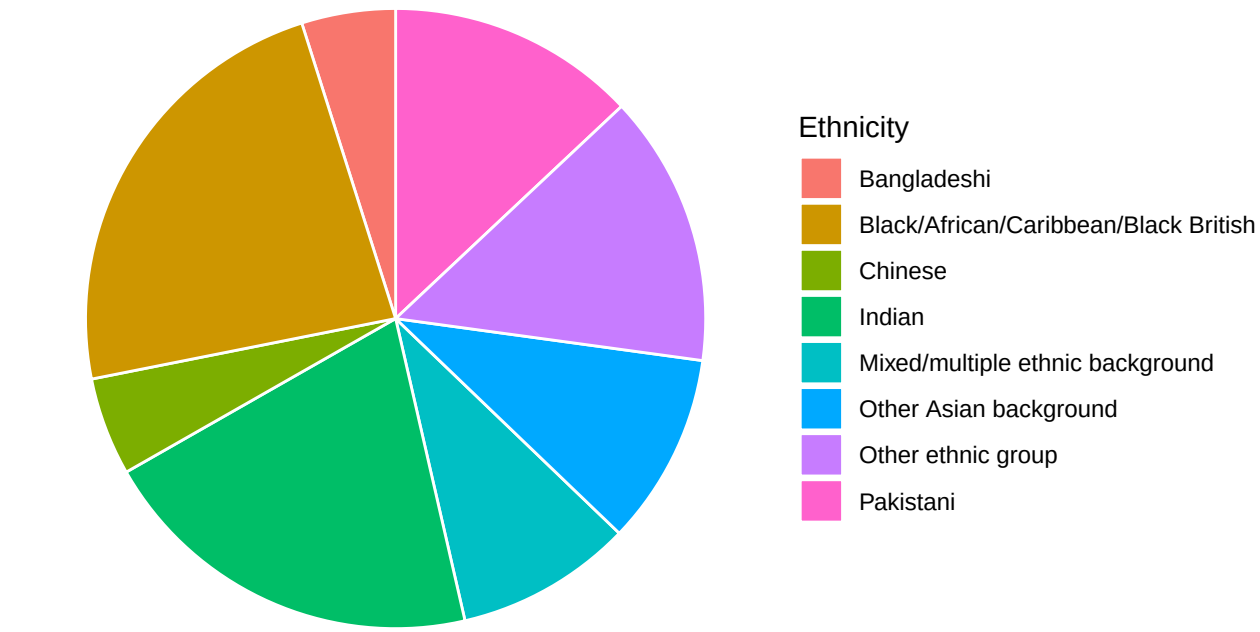
```
ethnic_other <- ethnic_pivot %>% filter(!Ethnicity=="White")
ethnic_other_mean <- ethnic_mean %>% filter(!Ethnicity=="White")
```

```
e <- ggplot(ethnic_other,aes(x=Years, y = Users, fill =Ethnicity)) + geom_col(position = "dodge2") + lab
ggsave("ethnic_other.png", e, width = 15, height = 10)
e
```



Pie Chart of ethnic_other

```
pie <- ggplot(ethnic_other_mean, aes(x="", y=mean_users, fill=Ethnicity)) +
  geom_bar(stat="identity", width=1, color="white") + labs(title = "Average Internet Usage by Ethnicity
  coord_polar("y", start=0) + theme_void()
ggsave("ethnic_other_pie.png", pie, width = 15, height = 10)
pie
```



Anova Test

```
ethnic_aov <- aov(Users ~ Ethnicity, data = ethnic_other)
summary(ethnic_aov)
```

```
##             Df  Sum Sq Mean Sq F value Pr(>F)
## Ethnicity    7 6146238  878034   104.7 <2e-16 ***
## Residuals   56  469830    8390
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Studying Linear Regression in White Ethnicity

```
ethnic_white <- ethnic_data %>% filter(Ethnicity == "White") %>% pivot_longer(-Ethnicity, names_to = "Y
ethnic_white_lm <- lm(Years ~ Users, data = ethnic_white)
summary(ethnic_white_lm)
```

```
##
## Call:
## lm(formula = Years ~ Users, data = ethnic_white)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.32339 -0.20119  0.01647  0.18701  0.28323
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.975e+03  1.683e+00 1173.72  < 2e-16 ***
## Users       1.096e-03  4.324e-05   25.34 2.49e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2545 on 6 degrees of freedom
## Multiple R-squared:  0.9907, Adjusted R-squared:  0.9892
## F-statistic: 642.4 on 1 and 6 DF,  p-value: 2.485e-07
```

```
ethnic_lm <- lm(formula = Users ~ Ethnicity, data = ethnic_pivot)
summary(ethnic_lm)
```

```
##
## Call:
## lm(formula = Users ~ Ethnicity, data = ethnic_pivot)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3316.0   -59.6     3.8    71.5  2963.0
##
## Coefficients:
##                                                Estimate Std. Error t value
## (Intercept)                                       244.1      264.0   0.925
## EthnicityBlack/African/Caribbean/Black British    914.2      373.3   2.449
## EthnicityChinese                                   11.0      373.3   0.029
## EthnicityIndian                                   770.5      373.3   2.064
## EthnicityMixed/multiple ethnic background         216.2      373.3   0.579
## EthnicityOther Asian background                   254.9      373.3   0.683
## EthnicityOther ethnic group                       462.2      373.3   1.238
## EthnicityPakistani                                404.0      373.3   1.082
## EthnicityWhite                                  38617.9      373.3 103.439
##                                                Pr(>|t|)
## (Intercept)                                      0.3586
## EthnicityBlack/African/Caribbean/Black British   0.0171 *
## EthnicityChinese                                 0.9766
## EthnicityIndian                                  0.0432 *
## EthnicityMixed/multiple ethnic background        0.5645
## EthnicityOther Asian background                  0.4973
## EthnicityOther ethnic group                      0.2203
## EthnicityPakistani                               0.2833
## EthnicityWhite                                   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 746.7 on 63 degrees of freedom
## Multiple R-squared:  0.9966, Adjusted R-squared:  0.9962
## F-statistic:  2333 on 8 and 63 DF,  p-value: < 2.2e-16
```

4. Economic Activity Data

```
eco_pivot <- eco_data %>% pivot_longer(-Activity, names_to = "Years", values_to = "Users")
#eco_pivot

eco_mean <- eco_data %>% mutate(mean_users = rowMeans(.[, years]))
eco_mean <- eco_mean %>% select(Activity,mean_users)
eco_mean
```

```
## # A tibble: 8 x 2
##   Activity                                      mean_users
##   <chr>                                              <dbl>
## 1 Employee                                         24978.
## 2 Self-employed                                     4172.
## 3 Government employment & training programmes         99.2
## 4 Unpaid family worker                                95.6
## 5 Unemployed                                        1888.
## 6 Student                                           2437.
## 7 Retired                                           5692.
## 8 Inactive                                          4518.
```

visualizing data

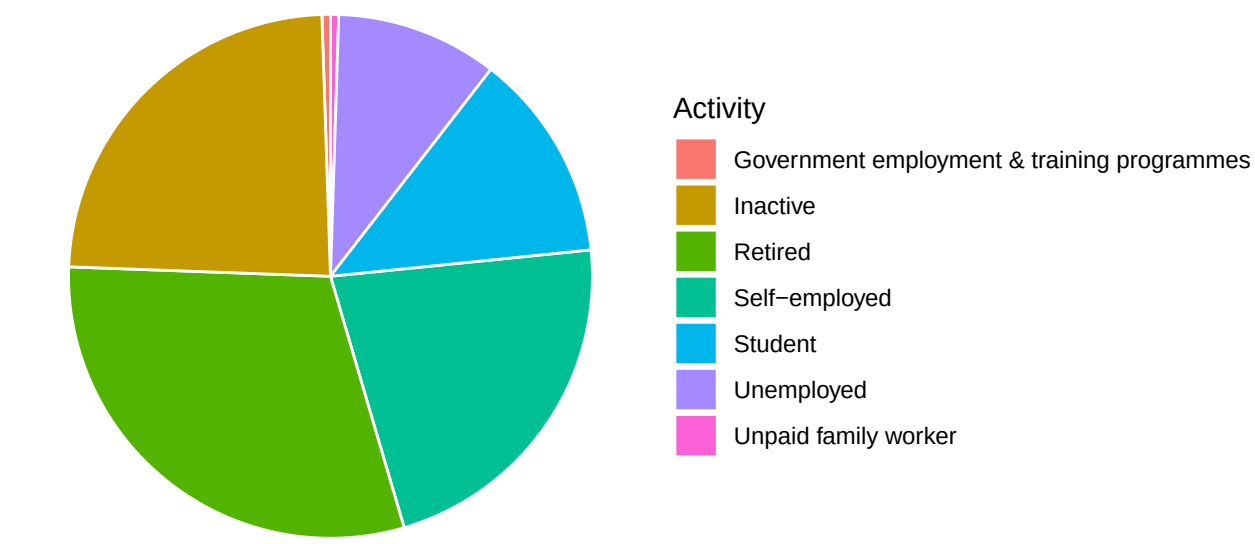
```
eco <- ggplot(eco_pivot,aes(x=Years, y = Users, fill = Activity)) + geom_col(position = "dodge2") + labs
ggsave("eco_data.png", eco, width = 15, height = 10)
eco
```

Removing employee group

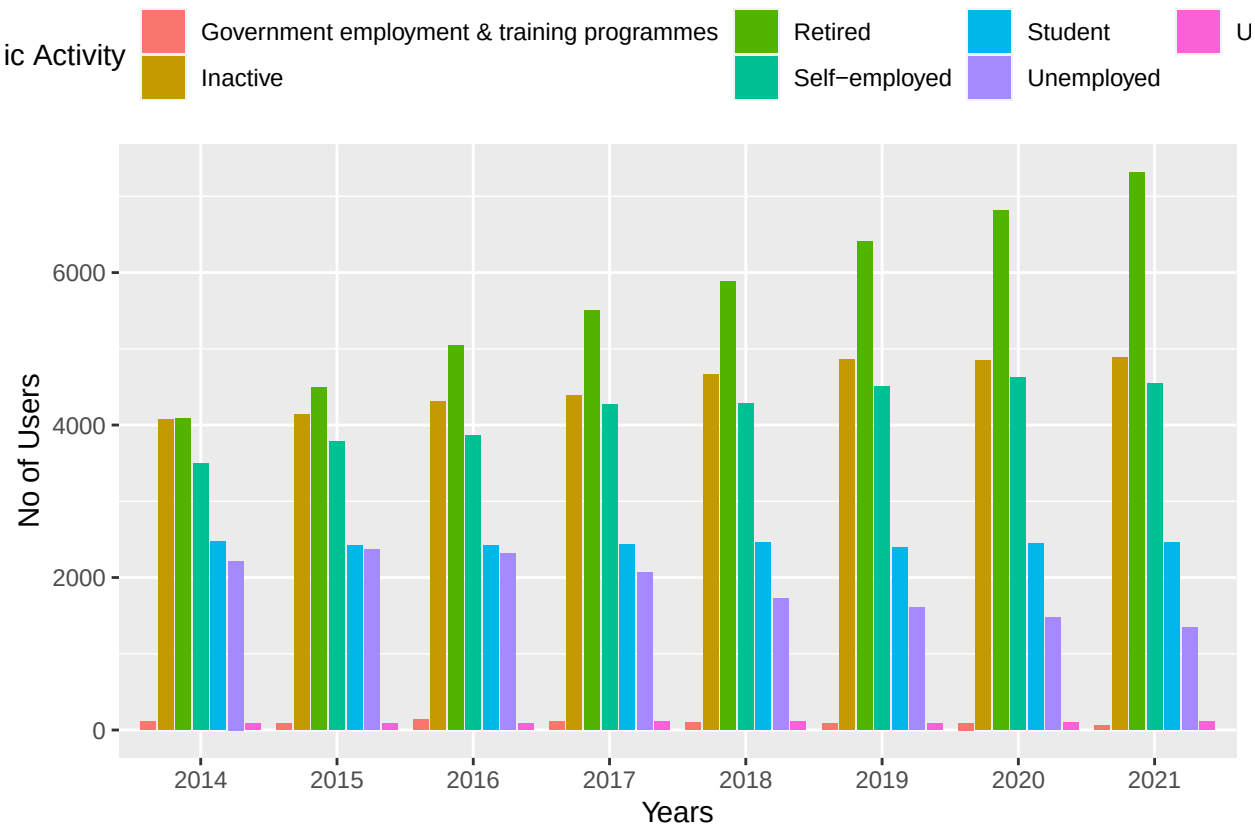
```
eco_other <- eco_pivot %>% filter(!Activity=="Employee")
eco_other_mean <- eco_mean %>% filter(!Activity=="Employee")
```

Visualizing without Employee

```
pie <- ggplot(eco_other_mean, aes(x="", y=mean_users, fill=Activity)) +
  geom_bar(stat="identity", width=1, color="white") + labs(title = "Average Internet Usage by Ethnicity
  coord_polar("y", start=0) + theme_void()
ggsave("ethnic_other_pie.png", pie, width = 15, height = 10)
pie
```




```
eco <- ggplot(eco_other,aes(x=Years, y = Users, fill = Activity)) + geom_col(position = "dodge2") + labs
ggsave("eco_other.png", eco, width = 15, height = 10)
eco
```



ANOVA TEST

```
eco_aov <- aov(Users ~ Activity, data = eco_other)
summary(eco_aov)
```

```
##              Df    Sum Sq  Mean Sq F value Pr(>F)
## Activity      6 229590292 38265049     156 <2e-16 ***
## Residuals    49  12018867   245283
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

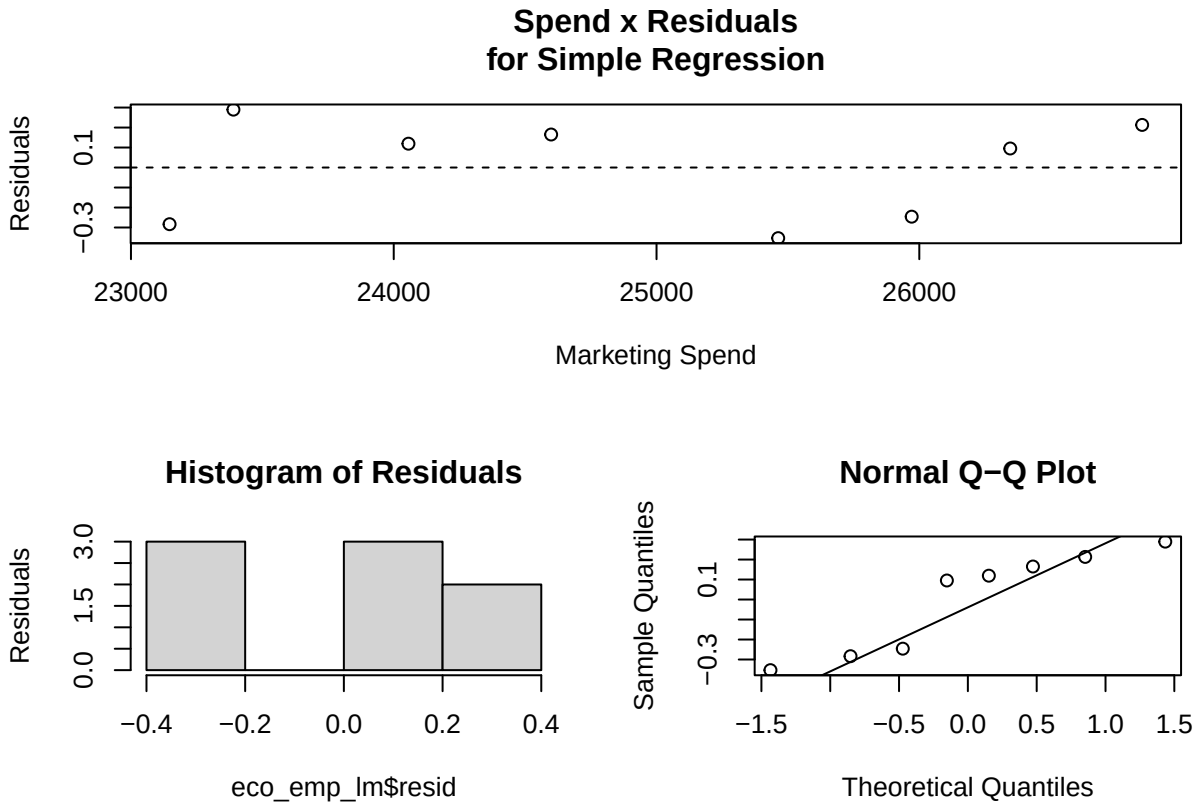
Linear Regression for Employee Group

```
eco_emp <- eco_data %>% filter(Activity == "Employee") %>% pivot_longer(-Activity, names_to = "Years", 
eco_emp_lm <- lm(Years ~ Users, data = eco_emp)
summary(eco_emp_lm)
```

```
##
## Call:
## lm(formula = Years ~ Users, data = eco_emp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3531 -0.2552  0.1073  0.1771  0.2896
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.974e+03  1.857e+00 1062.60  < 2e-16 ***
## Users       1.757e-03  7.426e-05   23.66 3.74e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2724 on 6 degrees of freedom
## Multiple R-squared:  0.9894, Adjusted R-squared:  0.9876
## F-statistic: 559.9 on 1 and 6 DF,  p-value: 3.739e-07
```

```r
layout(matrix(c(1,1,2,3),2,2,byrow=T))
#Spend x Residuals Plot
plot(eco_emp_lm$resid~eco_emp$Users[order(eco_emp$Users)],
 main="Spend x Residuals\nfor Simple Regression",
 xlab="Marketing Spend", ylab="Residuals")
abline(h=0,lty=2)
#Histogram of Residuals
hist(eco_emp_lm$resid, main="Histogram of Residuals",
 ylab="Residuals")
#Q-Q Plot
qqnorm(eco_emp_lm$resid)
qqline(eco_emp_lm$resid)
```



```r
eco_2 <- eco_data %>% filter(Activity == "Self-employed") %>% pivot_longer(-Activity, names_to = "Years"
eco_2_lm <- lm(Years ~ Users, data = eco_2)
summary(eco_2_lm)
```

```
##
## Call:
## lm(formula = Years ~ Users, data = eco_2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0387 -0.3383 -0.1161  0.2848  1.3781
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.994e+03  2.947e+00 676.458 7.04e-16 ***
## Users       5.716e-03  7.035e-04   8.125 0.000187 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7637 on 6 degrees of freedom
## Multiple R-squared:  0.9167, Adjusted R-squared:  0.9028
## F-statistic: 66.01 on 1 and 6 DF,  p-value: 0.0001867
```

Retired

```r
eco_3 <- eco_data %>% filter(Activity == "Retired") %>% pivot_longer(-Activity, names_to = "Years", valu
eco_3_lm <- lm(Years ~ Users, data = eco_3)
summary(eco_3_lm)
```

```
##
## Call:
## lm(formula = Years ~ Users, data = eco_3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.09998 -0.05137 -0.01844  0.07457  0.10158
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept) 2.005e+03  1.597e-01 12557.37  < 2e-16 ***
## Users       2.169e-03  2.758e-05    78.64 2.85e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08237 on 6 degrees of freedom
## Multiple R-squared:  0.999,  Adjusted R-squared:  0.9989
## F-statistic:  6184 on 1 and 6 DF,  p-value: 2.847e-10
```

Student

```r
eco_4 <- eco_data %>% filter(Activity == "Student") %>% pivot_longer(-Activity, names_to = "Years", valu
eco_4_lm <- lm(Years ~ Users, data = eco_4)
summary(eco_4_lm)
```

```
##
## Call:
## lm(formula = Years ~ Users, data = eco_4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3153 -1.8314  0.0375  1.5610  3.6225
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.031e+03  9.410e+01  21.586 6.45e-07 ***
## Users       -5.663e-03  3.861e-02  -0.147    0.888
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.641 on 6 degrees of freedom
## Multiple R-squared:  0.003573,   Adjusted R-squared:  -0.1625
## F-statistic: 0.02151 on 1 and 6 DF,  p-value: 0.8882
```

Inactive

```r
eco_5 <- eco_data %>% filter(Activity == "Inactive") %>% pivot_longer(-Activity, names_to = "Years", val
eco_5_lm <- lm(Years ~ Users, data = eco_5)
summary(eco_5_lm)
```

```
##
## Call:
## lm(formula = Years ~ Users, data = eco_5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.89894 -0.36406  0.08774  0.28343  0.85939
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.985e+03  3.081e+00  644.44 9.42e-16 ***
## Users       7.108e-03  6.803e-04   10.45 4.51e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6039 on 6 degrees of freedom
## Multiple R-squared:  0.9479, Adjusted R-squared:  0.9392
## F-statistic: 109.2 on 1 and 6 DF,  p-value: 4.509e-05
```

5. Disability Data

```r
all_disability <- disability_data[1:2,] %>% select(!AgeGroup)
all_disability
```

```
## # A tibble: 2 x 6
##   Disability                 `2017` `2018` `2019` `2020` `2021`
##   <chr>                       <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1 Equality Act Disabled1       7413   8038   8586   8984   9543
## 2 Not Equality Act Disabled2  36044  36633  37330  37758  38017
```

```r
disability_age <- tail(disability_data,-2)
disability_age
```

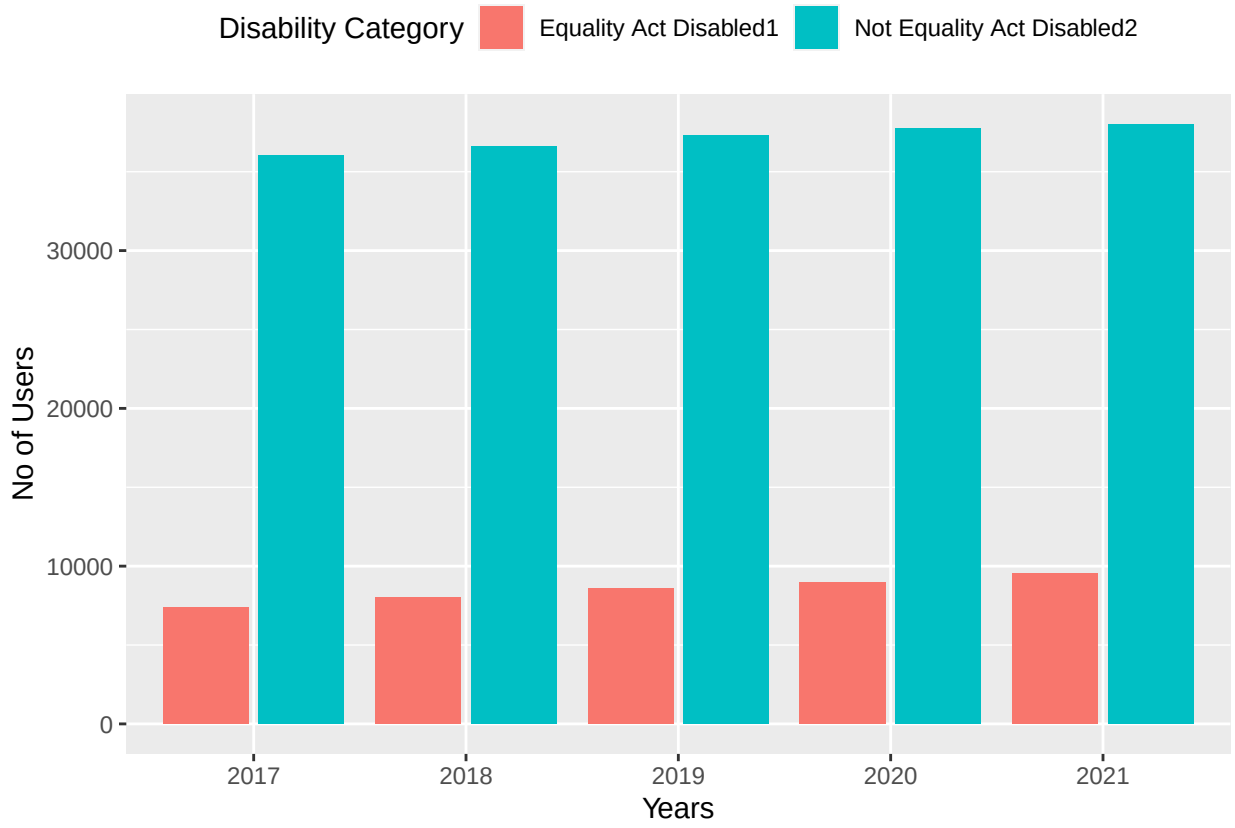
```
## # A tibble: 14 x 7
##    AgeGroup Disability                `2017` `2018` `2019` `2020` `2021`
##    <chr>    <chr>                      <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
##  1 16-24    Equality Act Disabled        697    752    824    834    824
##  2 16-24    Not Equality Act Disabled   6377   6403   6306   6202   6168
##  3 25-34    Equality Act Disabled        864    923   1026   1031   1144
##  4 25-34    Not Equality Act Disabled   7796   7660   7694   7785   7750
##  5 35-44    Equality Act Disabled       1091   1127   1185   1194   1207
##  6 35-44    Not Equality Act Disabled   6809   6926   6945   6924   6937
##  7 45-54    Equality Act Disabled       1421   1562   1591   1634   1732
##  8 45-54    Not Equality Act Disabled   6869   6936   7095   7169   7082
##  9 55-64    Equality Act Disabled       1472   1594   1651   1725   1845
## 10 55-64    Not Equality Act Disabled   4588   4767   4956   5163   5344
## 11 65-74    Equality Act Disabled       1198   1330   1477   1643   1721
## 12 65-74    Not Equality Act Disabled   2741   3060   3245   3388   3543
## 13 75+      Equality Act Disabled        672    751    834    923   1070
## 14 75+      Not Equality Act Disabled    862    881   1091   1128   1192
```

```r
all_dis_pivot <- all_disability %>% pivot_longer(-Disability,names_to = "Years", values_to = "Users")
all_dis_pivot
```

```
## # A tibble: 10 x 3
##    Disability                 Years Users
##    <chr>                      <chr> <dbl>
##  1 Equality Act Disabled1     2017   7413
##  2 Equality Act Disabled1     2018   8038
##  3 Equality Act Disabled1     2019   8586
##  4 Equality Act Disabled1     2020   8984
##  5 Equality Act Disabled1     2021   9543
##  6 Not Equality Act Disabled2 2017  36044
##  7 Not Equality Act Disabled2 2018  36633
##  8 Not Equality Act Disabled2 2019  37330
##  9 Not Equality Act Disabled2 2020  37758
## 10 Not Equality Act Disabled2 2021  38017
```

```r
disability <- as.vector(all_dis_pivot$Disability)

p<- ggplot(all_dis_pivot,aes(x=Years,y=Users,fill=disability)) + geom_col(position = "dodge2") + labs(x=
ggsave("dis_all.png", p, width = 15, height = 10)
p
```




```
dis_pivot <- disability_age %>% filter(Disability=="Equality Act Disabled") %>% select(!Disability) %>%
dis_pivot
```

```
## # A tibble: 35 x 3
##    AgeGroup Years Users
##    <chr>    <chr> <dbl>
##  1 16-24    2017    697
##  2 16-24    2018    752
##  3 16-24    2019    824
##  4 16-24    2020    834
##  5 16-24    2021    824
##  6 25-34    2017    864
##  7 25-34    2018    923
##  8 25-34    2019   1026
##  9 25-34    2020   1031
## 10 25-34    2021   1144
## # ... with 25 more rows
```

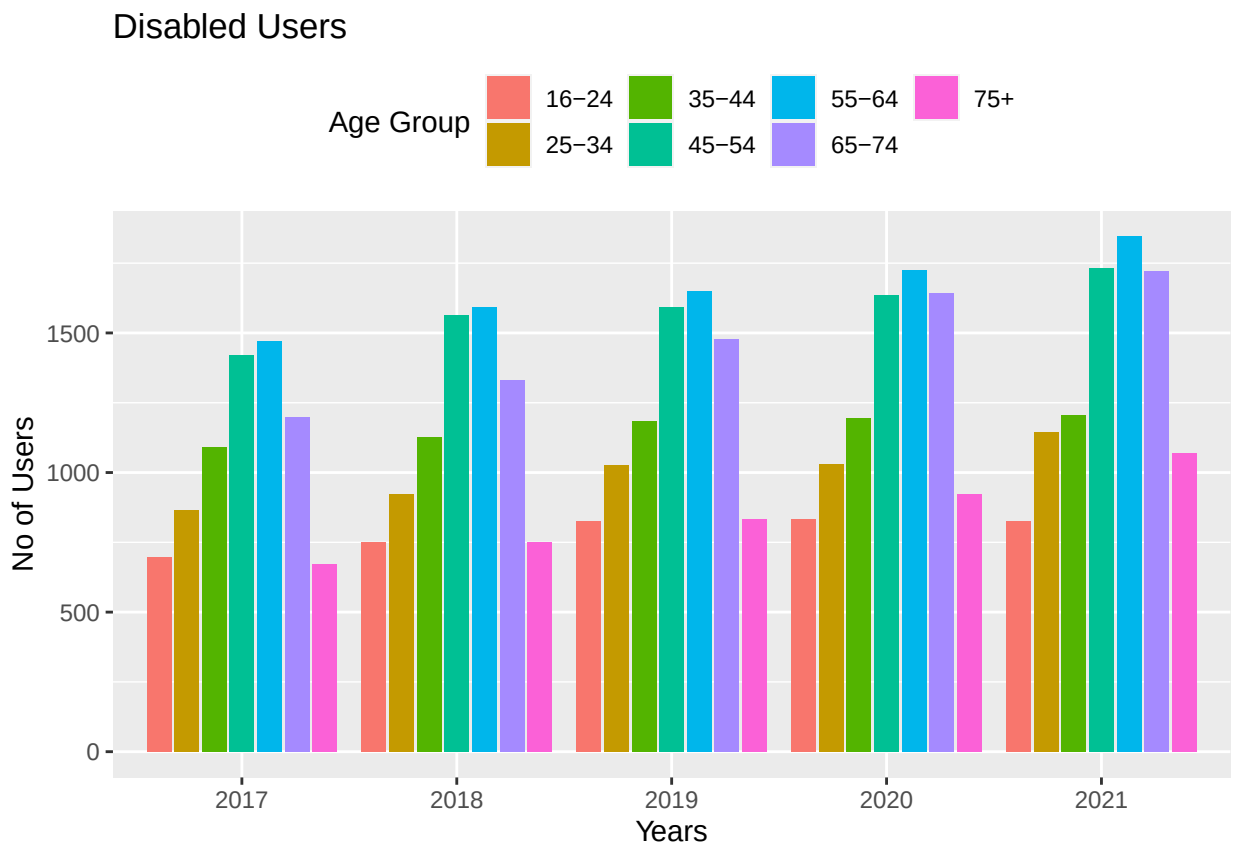
```
dis_non_pivot <- disability_age %>% filter(Disability == "Not Equality Act Disabled") %>% select(!Disabi
  pivot_longer(-AgeGroup, names_to = "Years", values_to = "Users")
dis_non_pivot
```

```
## # A tibble: 35 x 3
##    AgeGroup Years Users
```

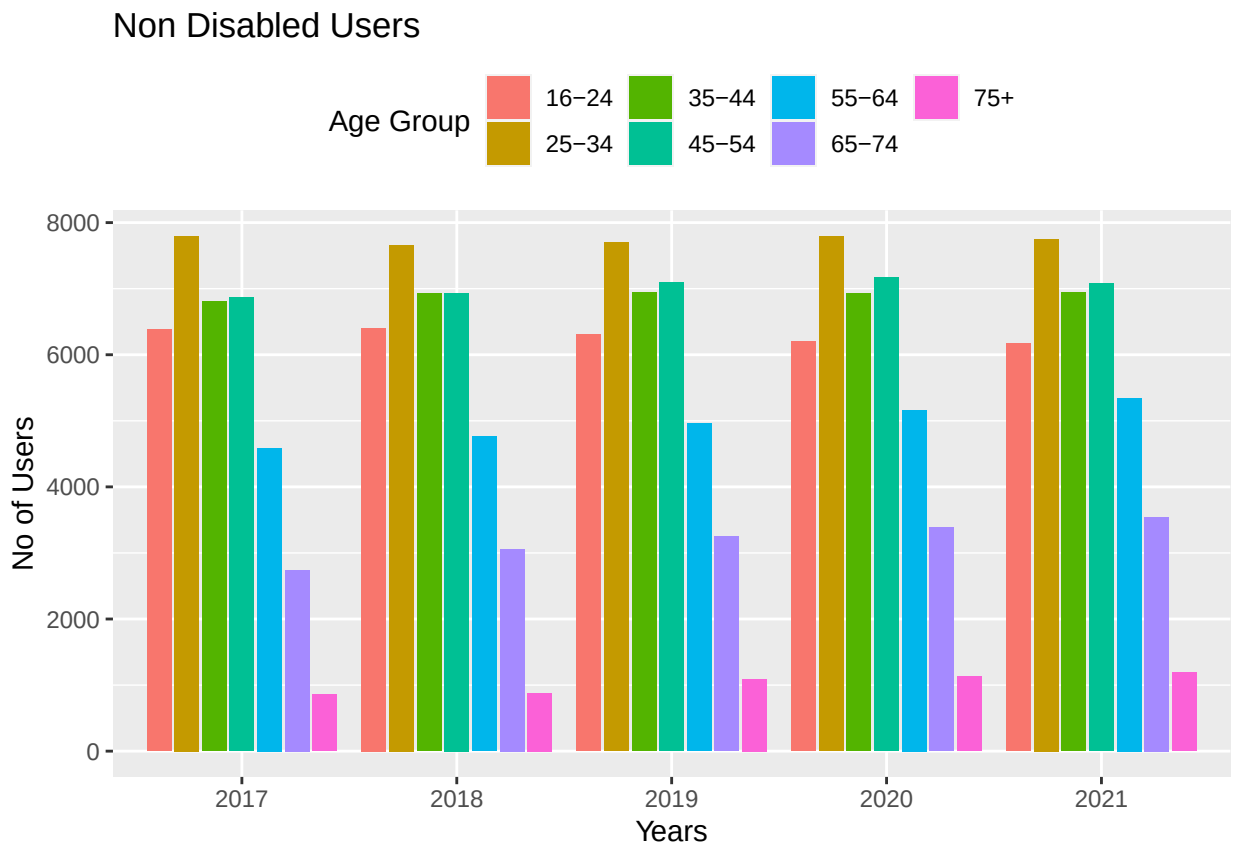
```
##    <chr>    <chr> <dbl>
##  1 16-24    2017   6377
##  2 16-24    2018   6403
##  3 16-24    2019   6306
##  4 16-24    2020   6202
##  5 16-24    2021   6168
##  6 25-34    2017   7796
##  7 25-34    2018   7660
##  8 25-34    2019   7694
##  9 25-34    2020   7785
## 10 25-34    2021   7750
## # ... with 25 more rows
```

```
yes<- ggplot(dis_pivot,aes(x=Years,y=Users,fill=AgeGroup)) +geom_col(position="dodge2") + labs(title="D
no <- ggplot(dis_non_pivot,aes(x=Years,y=Users,fill=AgeGroup)) +geom_col(position="dodge2") + labs(titl

yes
```




```
no
```



Mean Data

```
years2 <- c("2017","2018","2019","2020","2021")
dis_mean <- disability_age %>% filter(Disability=="Equality Act Disabled") %>%
  select(!Disability) %>% mutate(mean_users = rowMeans(.[, years2]))
dis_mean <- dis_mean %>% select(AgeGroup,mean_users)
dis_mean
```

```
## # A tibble: 7 x 2
##   AgeGroup mean_users
##   <chr>         <dbl>
## 1 16-24          786.
## 2 25-34          998.
## 3 35-44         1161.
## 4 45-54         1588
## 5 55-64         1657.
## 6 65-74         1474.
## 7 75+            850
```

```
dis_non_mean <- disability_age %>% filter(Disability == "Not Equality Act Disabled") %>%
  select(!Disability) %>% mutate(mean_users = rowMeans(.[,years2]))
dis_non_mean <- dis_non_mean %>% select(AgeGroup, mean_users)
dis_non_mean
```

```
## # A tibble: 7 x 2
##   AgeGroup mean_users
```

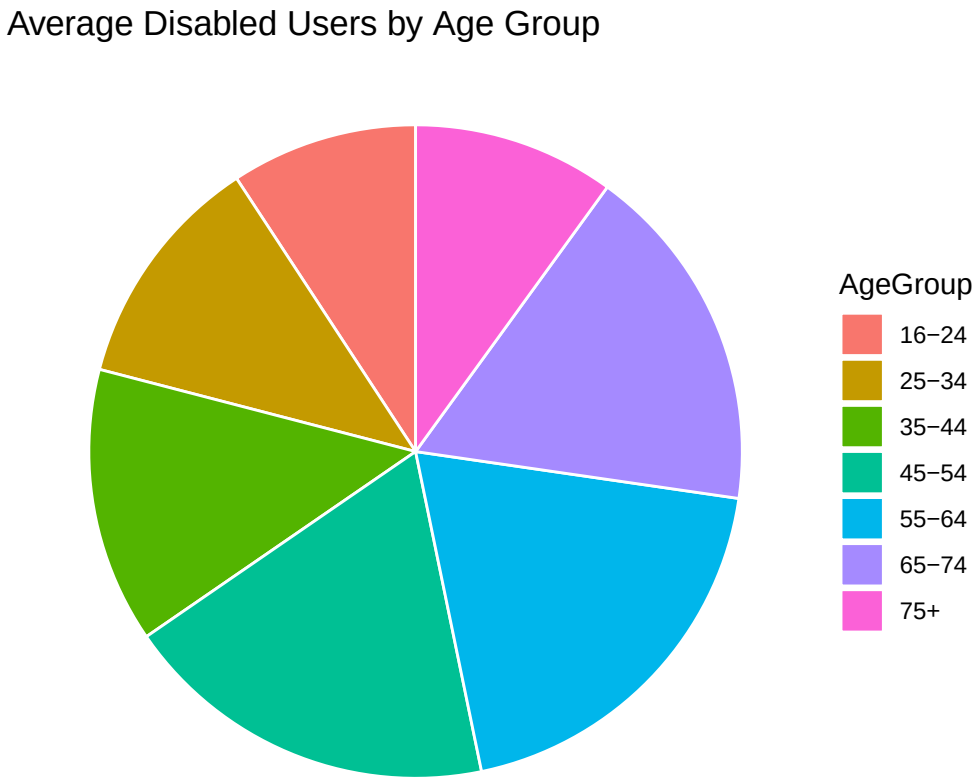
```
##   <chr>         <dbl>
## 1 16-24         6291.
## 2 25-34         7737
## 3 35-44         6908.
## 4 45-54         7030.
## 5 55-64         4964.
## 6 65-74         3195.
## 7 75+           1031.
```

```
dis_both_mean <- disability_age %>% mutate(mean_users = rowMeans(.[,years2])) %>% select(AgeGroup,Disabi
dis_both_mean
```

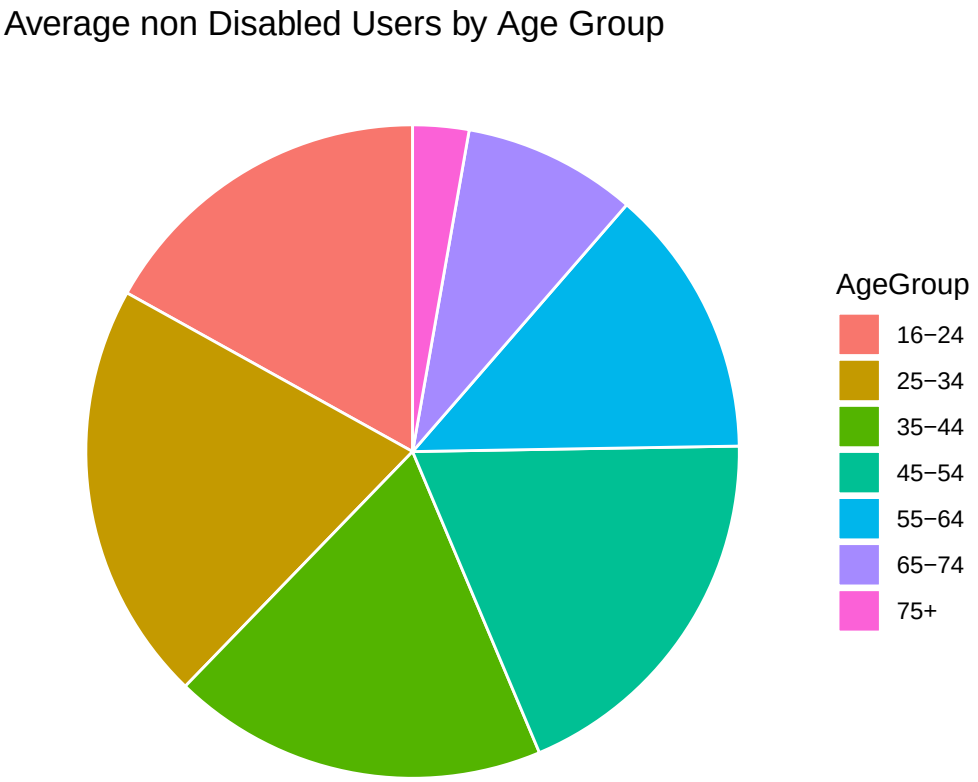
```
## # A tibble: 14 x 3
##    AgeGroup Disability                mean_users
##    <chr>    <chr>                          <dbl>
##  1 16-24    Equality Act Disabled           786.
##  2 16-24    Not Equality Act Disabled      6291.
##  3 25-34    Equality Act Disabled           998.
##  4 25-34    Not Equality Act Disabled      7737
##  5 35-44    Equality Act Disabled          1161.
##  6 35-44    Not Equality Act Disabled      6908.
##  7 45-54    Equality Act Disabled          1588
##  8 45-54    Not Equality Act Disabled      7030.
##  9 55-64    Equality Act Disabled          1657.
## 10 55-64    Not Equality Act Disabled      4964.
## 11 65-74    Equality Act Disabled          1474.
## 12 65-74    Not Equality Act Disabled      3195.
## 13 75+      Equality Act Disabled           850
## 14 75+      Not Equality Act Disabled      1031.
```

PI Charts for both

```
pie1 <- ggplot(dis_mean, aes(x="", y=mean_users, fill=AgeGroup)) +
  geom_bar(stat="identity", width=1, color="white") + labs(title = "Average Disabled Users by Age Group
  coord_polar("y", start=0) + theme_void()
ggsave("dis_mean_pie.png", pie1, width = 15, height = 10)
pie1
```

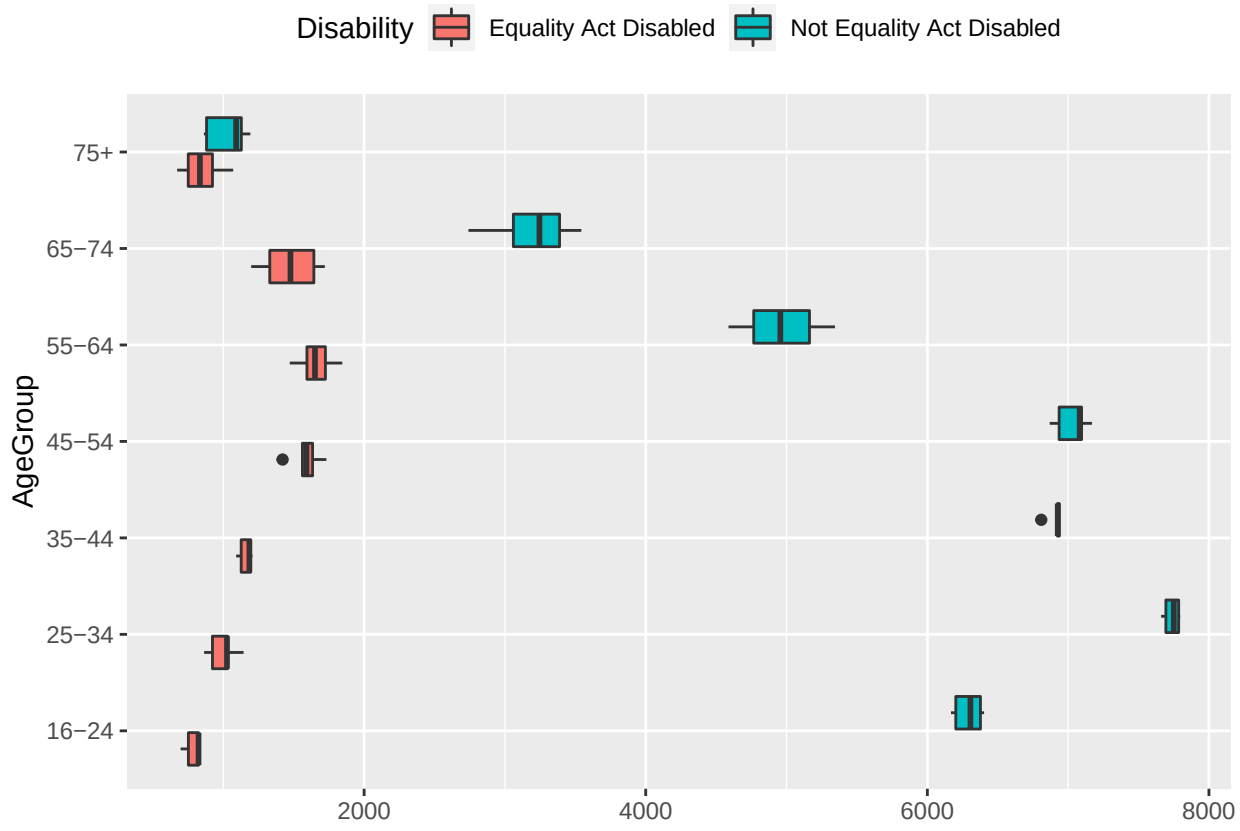



```
pie2 <- ggplot(dis_non_mean, aes(x="", y=mean_users, fill=AgeGroup)) +
  geom_bar(stat="identity", width=1, color="white") + labs(title = "Average non Disabled Users by Age G
  coord_polar("y", start=0) + theme_void()
ggsave("dis_non_mean_pie.png", pie2, width = 15, height = 10)
pie2
```



Boxplot

```
dis_pivot <- gather(disability_age, "Years", "Users", 3:7 )
dis <- ggplot(dis_pivot,aes(x=Users, y = AgeGroup, fill=Disability)) + geom_boxplot() +
  labs(x=NULL) + theme(legend.position = "top")
ggsave("dis_box.png", dis, width = 15, height = 10)
dis
```




```
t.test(data = all_dis_pivot, Users ~ Disability)
```

```
##
##  Welch Two Sample t-test
##
## data:  Users by Disability
## t = -55.29, df = 7.9982, p-value = 1.276e-11
## alternative hypothesis: true difference in means between group Equality Act Disabled1 and group Not
## 95 percent confidence interval:
##  -29838.3 -27448.9
## sample estimates:
##     mean in group Equality Act Disabled1
##                                   8512.8
## mean in group Not Equality Act Disabled2
##                                  37156.4
```

```
#dis_t <- t.test(Users ~ Disability, data = dis_pivot)
#summary(dis_t)

dis_lm <- lm(formula = Users ~ Years + Disability, data = all_dis_pivot)
summary(dis_lm)
```

```
##
## Call:
```

```
## lm(formula = Users ~ Years + Disability, data = all_dis_pivot)
##
## Residuals:
##     1     2     3     4     5     6     7     8     9    10
##   6.3  24.3 -50.2 -65.2  84.8  -6.3 -24.3  50.2  65.2 -84.8
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           7406.70      66.16 111.944 3.82e-08 ***
## Years2018                              607.00      85.42   7.106 0.002072 **
## Years2019                             1229.50      85.42  14.394 0.000135 ***
## Years2020                             1642.50      85.42  19.229 4.31e-05 ***
## Years2021                             2051.50      85.42  24.017 1.78e-05 ***
## DisabilityNot Equality Act Disabled2 28643.60      54.02 530.213 7.59e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 85.42 on 4 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 5.637e+04 on 5 and 4 DF,  p-value: 8.811e-10
```

```
png("age_data.png", height=50 * nrow(age_data), width=50 * ncol(age_data), bg="white")
p<-tableGrob(age_data)
grid.arrange(p)
dev.off()
```

```
## pdf
##   2
```

```
png("sex_data.png", height=50 * nrow(age_data), width=50 * ncol(age_data), bg="white")
p<-tableGrob(sex_data)
grid.arrange(p)
dev.off()
```

```
## pdf
##   2
```

```
png("age_sex.png", height=55 * nrow(age_data), width=80 * ncol(age_data), bg="white")
p<-tableGrob(age_sex)
grid.arrange(p)
dev.off()
```

```
## pdf
##   2
```

```
png("eco_data.png", height=55 * nrow(age_data), width=80 * ncol(age_data), bg="white")
p<-tableGrob(eco_data)
grid.arrange(p)
dev.off()
```

```
## pdf
##   2
```

```
png("ethnic_data.png", height=60 * nrow(age_data), width=80 * ncol(age_data), bg="white")
p<-tableGrob(ethnic_data)
grid.arrange(p)
dev.off()
```

```
## pdf
##   2
```

```
png("disability_data.png", height=55 * nrow(age_data), width=80 * ncol(age_data), bg="white")
p<-tableGrob(disability_age)
grid.arrange(p)
dev.off()
```

```
## pdf
##   2
```

```
png("age_pivot.png", height=55 * nrow(age_pivot), width=80 * ncol(age_pivot), bg="white")
p<-tableGrob(age_pivot)
grid.arrange(p)
dev.off()
```

```
## pdf
##   2
```

6. Age & Sex Data

```
all_sex <- age_sex[1:2,] %>% select(!AgeGroup)
all_sex
```

```
## # A tibble: 2 x 9
##   Sex   `2014` `2015` `2016` `2017` `2018` `2019` `2020` `2021`
##   <chr>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1 Men    20039  20564  21242  21814  22229  22812  23200  23606
## 2 Women  19645  20200  21001  21643  22442  23105  23542  23954
```

```
all_sex_pivot <- all_sex %>% pivot_longer(-Sex,names_to = "Years", values_to = "Users")
all_sex_pivot
```

```
## # A tibble: 16 x 3
##    Sex   Years Users
##    <chr> <chr> <dbl>
##  1 Men   2014  20039
##  2 Men   2015  20564
##  3 Men   2016  21242
##  4 Men   2017  21814
##  5 Men   2018  22229
##  6 Men   2019  22812
##  7 Men   2020  23200
##  8 Men   2021  23606
##  9 Women 2014  19645
```

```
## 10 Women 2015  20200
## 11 Women 2016  21001
## 12 Women 2017  21643
## 13 Women 2018  22442
## 14 Women 2019  23105
## 15 Women 2020  23542
## 16 Women 2021  23954
```

```
agesex_data <- tail(age_sex,-2)
agesex_data
```

```
## # A tibble: 14 x 10
##    AgeGroup Sex   `2014` `2015` `2016` `2017` `2018` `2019` `2020` `2021`
##    <chr>    <chr>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
##  1 16-24    Men     3643   3610   3593   3590   3638   3622   3594   3561
##  2 16-24    Women   3503   3477   3482   3484   3517   3507   3443   3431
##  3 25-34    Men     3989   4136   4272   4408   4276   4340   4402   4454
##  4 25-34    Women   3914   4026   4186   4252   4307   4380   4413   4440
##  5 35-44    Men     4010   3969   3939   3925   3975   4018   4011   4025
##  6 35-44    Women   4064   4016   4013   3976   4078   4111   4107   4120
##  7 45-54    Men     3661   3803   3945   4058   4182   4270   4314   4310
##  8 45-54    Women   3770   3891   4059   4232   4315   4416   4489   4504
##  9 55-64    Men     2710   2807   2877   3004   3118   3240   3375   3516
## 10 55-64    Women   2724   2817   2944   3056   3244   3367   3513   3673
## 11 65-74    Men     1472   1652   1847   1984   2183   2323   2471   2580
## 12 65-74    Women   1327   1501   1715   1955   2207   2398   2560   2684
## 13 75+      Men      554    587    769    846    858    998   1033   1159
## 14 75+      Women    344    471    603    688    774    926   1017   1103
```

```
agesex_pivot <- gather(agesex_data, "Years", "Users", 3:10)
agesex_pivot
```

```
## # A tibble: 112 x 4
##    AgeGroup Sex   Years Users
##    <chr>    <chr> <chr> <dbl>
##  1 16-24    Men   2014   3643
##  2 16-24    Women 2014   3503
##  3 25-34    Men   2014   3989
##  4 25-34    Women 2014   3914
##  5 35-44    Men   2014   4010
##  6 35-44    Women 2014   4064
##  7 45-54    Men   2014   3661
##  8 45-54    Women 2014   3770
##  9 55-64    Men   2014   2710
## 10 55-64    Women 2014   2724
## # ... with 102 more rows
```

```
agesex_male <- agesex_data %>% filter(Sex=="Men") %>% select(!Sex) %>% pivot_longer(-AgeGroup, names_to=
agesex_male
```

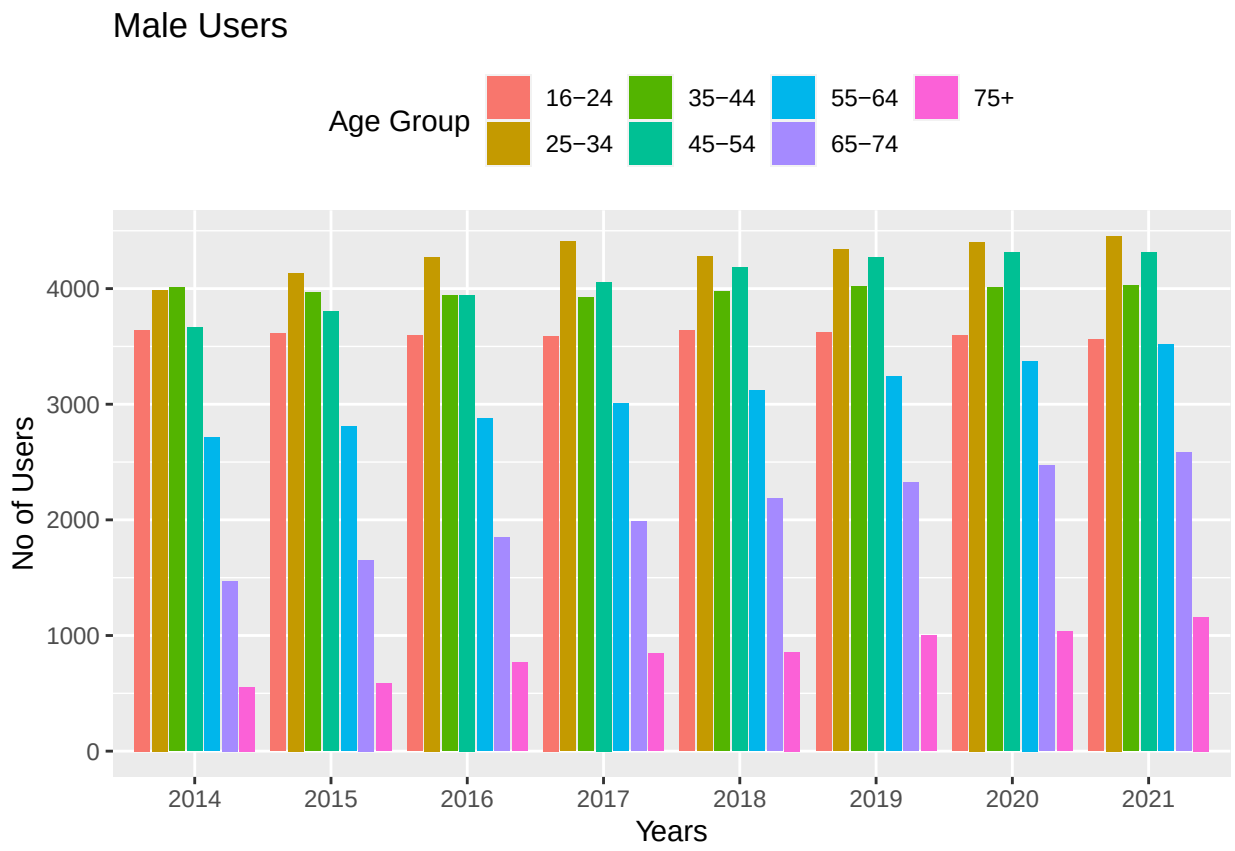
```
## # A tibble: 56 x 3
##    AgeGroup Years Users
```

```
##    <chr>    <chr> <dbl>
##  1 16-24    2014   3643
##  2 16-24    2015   3610
##  3 16-24    2016   3593
##  4 16-24    2017   3590
##  5 16-24    2018   3638
##  6 16-24    2019   3622
##  7 16-24    2020   3594
##  8 16-24    2021   3561
##  9 25-34    2014   3989
## 10 25-34    2015   4136
## # ... with 46 more rows
```

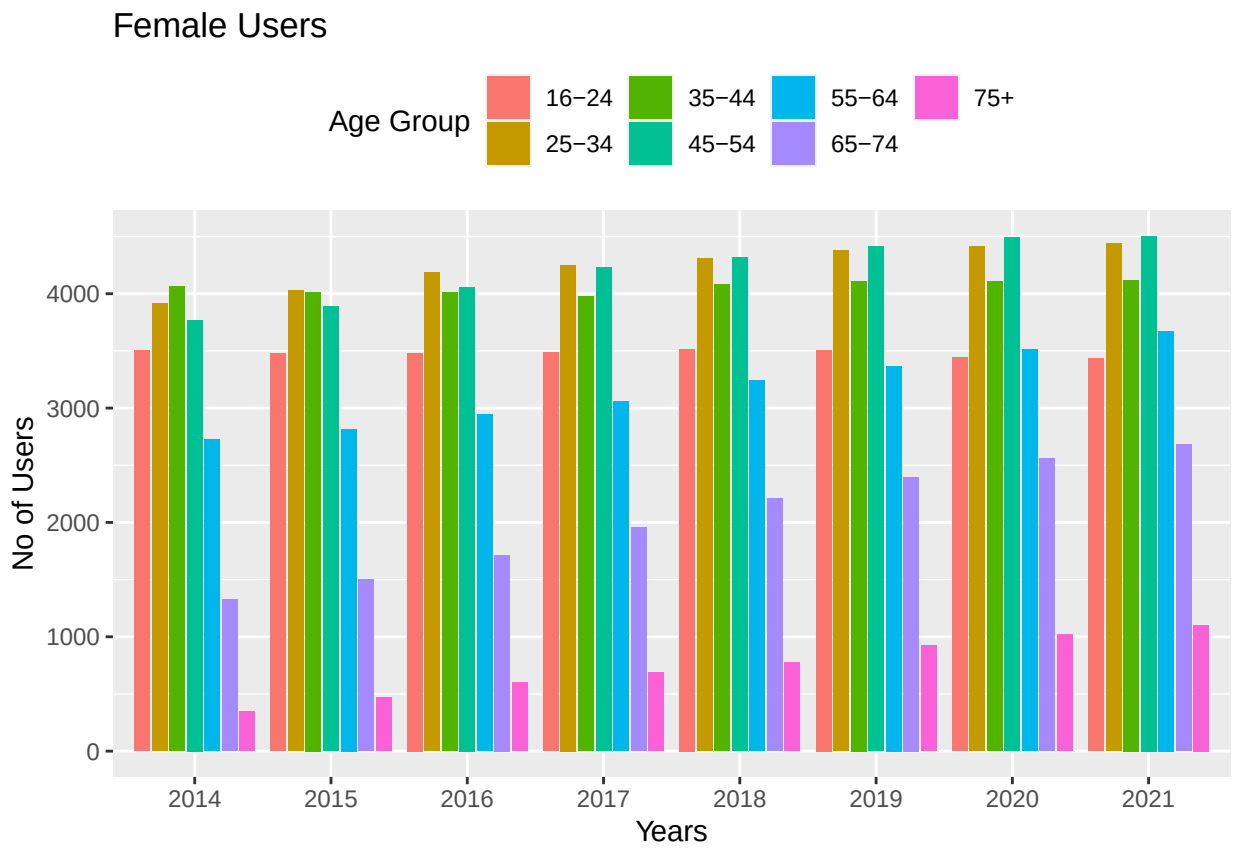
```
agesex_female <- agesex_data %>% filter(Sex == "Women") %>% select(!Sex) %>%
  pivot_longer(-AgeGroup, names_to = "Years", values_to = "Users")
agesex_female
```

```
## # A tibble: 56 x 3
##    AgeGroup Years Users
##    <chr>    <chr> <dbl>
##  1 16-24    2014   3503
##  2 16-24    2015   3477
##  3 16-24    2016   3482
##  4 16-24    2017   3484
##  5 16-24    2018   3517
##  6 16-24    2019   3507
##  7 16-24    2020   3443
##  8 16-24    2021   3431
##  9 25-34    2014   3914
## 10 25-34    2015   4026
## # ... with 46 more rows
```

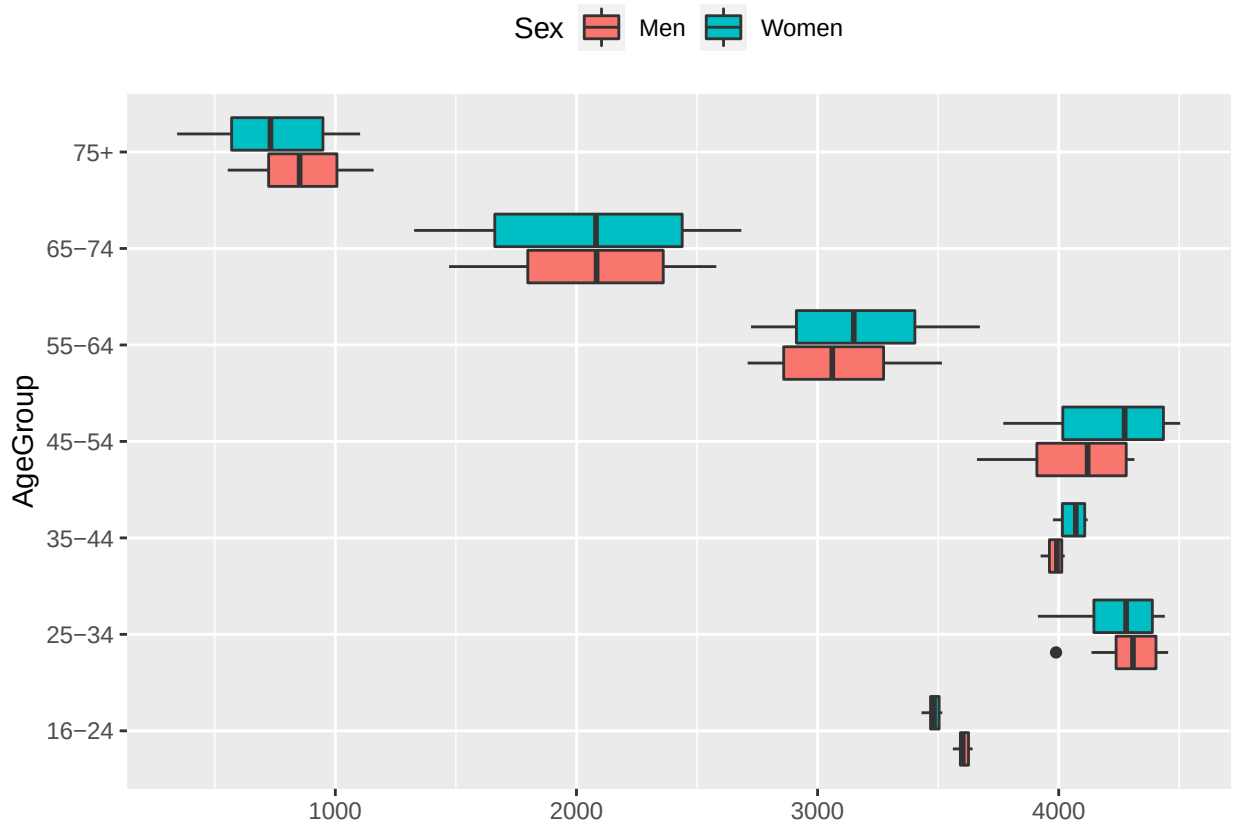
```
men<- ggplot(agesex_male,aes(x=Years,y=Users,fill=AgeGroup)) +geom_col(position="dodge2") + labs(title=
women<- ggplot(agesex_female,aes(x=Years,y=Users,fill=AgeGroup)) +geom_col(position="dodge2") + labs(ti
ggsave("men_pie.png", men, width = 15, height = 10)
ggsave("women_pie.png", women, width = 15, height = 10)
men
```


Male Users
Age Group
16−24
25−34
35−44
45−54
55−64
65−74
75+
No of Users
0
1000
2000
3000
4000
2014
2015
2016
2017
2018
2019
2020
2021
Years

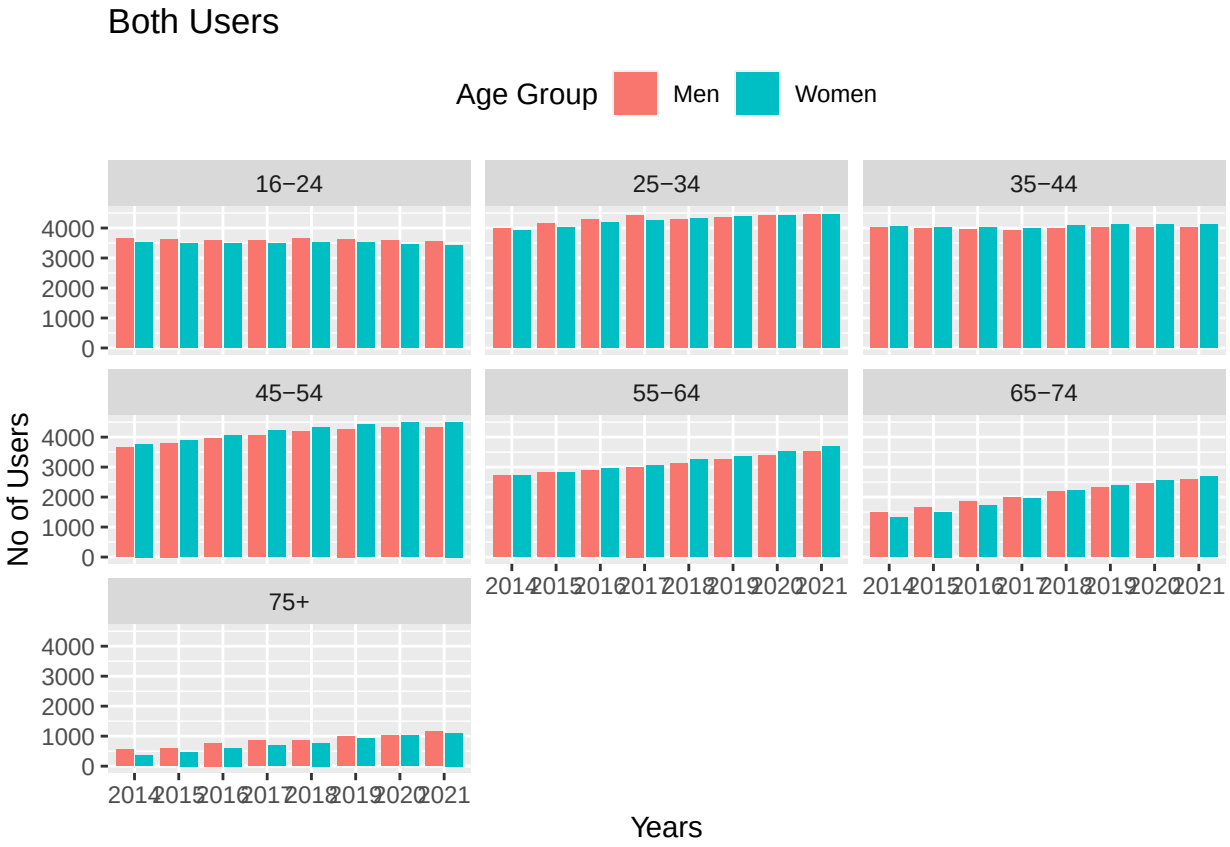
```
women
```




```
box <- ggplot(agesex_pivot,aes(x=Users, y = AgeGroup, fill=Sex)) + geom_boxplot() +
  labs(x=NULL) + theme(legend.position = "top")
ggsave("agesex_box.png", box, width = 15, height = 10)
box
```




```
both <- ggplot(agesex_pivot,aes(x=Years,y=Users,fill=Sex)) +geom_col(position="dodge2") + labs(title="B
ggsave("both_bar.png", both, width = 15, height = 10)
both
```




```
t.test(data = all_sex_pivot, Users ~ Sex)
```

```
##
##  Welch Two Sample t-test
##
## data:  Users by Sex
## t = -0.0045387, df = 13.361, p-value = 0.9964
## alternative hypothesis: true difference in means between group Men and group Women is not equal to 0
## 95 percent confidence interval:
##  -1545.968  1539.468
## sample estimates:
##   mean in group Men mean in group Women
##            21938.25            21941.50
```

```
#dis_t <- t.test(Users ~ Disability, data = dis_pivot)
#summary(dis_t)

ags_lm <- lm(formula = Users ~ Years + Sex, data = all_sex_pivot)
summary(ags_lm)
```

```
##
## Call:
## lm(formula = Users ~ Years + Sex, data = all_sex_pivot)
##
```

```
## Residuals:
##    Min     1Q Median     3Q    Max
## -198.6 -151.0    0.0  151.0  198.6
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19840.38     172.93 114.734 1.01e-12 ***
## Years2015     540.00     230.57   2.342  0.05169 .
## Years2016    1279.50     230.57   5.549  0.00086 ***
## Years2017    1886.50     230.57   8.182 7.89e-05 ***
## Years2018    2493.50     230.57  10.815 1.27e-05 ***
## Years2019    3116.50     230.57  13.517 2.85e-06 ***
## Years2020    3529.00     230.57  15.306 1.22e-06 ***
## Years2021    3938.00     230.57  17.080 5.79e-07 ***
## SexWomen        3.25     115.28   0.028  0.97830
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 230.6 on 7 degrees of freedom
## Multiple R-squared:  0.987,  Adjusted R-squared:  0.9722
## F-statistic: 66.64 on 8 and 7 DF,  p-value: 6.445e-06
```

```
ags_lm_age <- lm(formula = Users ~ AgeGroup + Sex, data = agesex_pivot)
summary(ags_lm_age)
```

```
##
## Call:
## lm(formula = Users ~ AgeGroup + Sex, data = agesex_pivot)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -726.94 -109.08   -0.56  133.69  630.06
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3543.19      67.79  52.269  < 2e-16 ***
## AgeGroup25-34   718.75      89.67   8.015 1.71e-12 ***
## AgeGroup35-44   478.88      89.67   5.340 5.49e-07 ***
## AgeGroup45-54   595.25      89.67   6.638 1.48e-09 ***
## AgeGroup55-64  -419.38      89.67  -4.677 8.80e-06 ***
## AgeGroup65-74 -1489.75      89.67 -16.613  < 2e-16 ***
## AgeGroup75+   -2747.81      89.67 -30.642  < 2e-16 ***
## SexWomen          0.50      47.93   0.010    0.992
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 253.6 on 104 degrees of freedom
## Multiple R-squared:  0.9594, Adjusted R-squared:  0.9566
## F-statistic: 350.8 on 7 and 104 DF,  p-value: < 2.2e-16
```

```
ags_aov <- aov(Users ~ AgeGroup + Sex, data = agesex_pivot)
summary(ags_aov)
```

```
##              Df   Sum Sq  Mean Sq F value Pr(>F)
## AgeGroup      6 1.58e+08 26330968   409.3 <2e-16 ***
## Sex           1 7.00e+00        7     0.0  0.992
## Residuals   104 6.69e+06    64331
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```