Project 512 Spring 2020

Vid Chan

4/18/2020

```
mergedata <- read.csv("/Users/vidarithchan/Desktop/MSAnalytics/ANLY 512/merged.csv")

colnames(mergedata)
```

```
##  [1] "Country"              "Year"              "Max_Partners"
##  [4] "GDP_per_unit_CO2"     "PPP_Conv_Rate"     "PPP_Share_GDP"
##  [7] "Imports_PC"           "Exports_PC"        "Govt_Revenue"
## [10] "gdp_per_cap"          "agri_perc_gdp"     "agg.empl.agri.perc"
## [13] "rural.pop.perc"       "pop.tot"           "mobilesub_per100peeps"
## [16] "intl_tourist_arrival" "total_life_exp"    "life_expectancy_fe"
## [19] "life_exp_male"        "trade_perGDP"
```

```
nrow(mergedata)
```

```
## [1] 2673
```

```
ncol(mergedata)
```

```
## [1] 20
```

```
# find missing-value count in each column
# confirming if Nicole has already cleaned the data
summary(mergedata)
```

```
##       Country       Year         Max_Partners   GDP_per_unit_CO2
##  Albania  : 27   Min.   :1990   Min.   : 45.0   Min.   : 0.5535
##  Algeria  : 27   1st Qu.:1996   1st Qu.:136.0   1st Qu.: 2.9293
##  Argentina: 27   Median :2003   Median :172.0   Median : 4.3336
##  Armenia  : 27   Mean   :2003   Mean   :168.3   Mean   : 5.1342
##  Australia: 27   3rd Qu.:2010   3rd Qu.:205.0   3rd Qu.: 6.2800
##  Austria  : 27   Max.   :2016   Max.   :235.0   Max.   :39.4786
##  (Other)  :2511
##  PPP_Conv_Rate     PPP_Share_GDP      Imports_PC        Exports_PC
##  Min.   :  0.001   Min.   : 0.0080   Min.   :-66.872   Min.   :-62.400
```

```
##    1st Qu.:    0.672    1st Qu.: 0.0570    1st Qu.:    0.732    1st Qu.:    0.754
##    Median :    1.693    Median : 0.2270    Median :    6.024    Median :    5.374
##    Mean   :   86.197    Mean   : 0.9433    Mean   :    6.592    Mean   :    5.791
##    3rd Qu.:   15.440    3rd Qu.: 0.6810    3rd Qu.:   12.400    3rd Qu.:   10.526
##    Max.   : 4085.960    Max.   :21.7800    Max.   : 112.650    Max.   :   93.959
##
##     Govt_Revenue         gdp_per_cap        agri_perc_gdp       agg.empl.agri.perc
##    Min.   :-29.88300    Min.   :   164.3    Min.   : 0.214    Min.   : 0.125
##    1st Qu.: -2.01900    1st Qu.:  2425.6    1st Qu.: 2.672    1st Qu.: 5.337
##    Median : -0.16800    Median :  7019.2    Median : 6.467    Median :15.695
##    Mean   : -0.01627    Mean   : 16052.6    Mean   : 9.791    Mean   :22.405
##    3rd Qu.:  1.96500    3rd Qu.: 24495.7    3rd Qu.:13.787    3rd Qu.:36.700
##    Max.   : 36.01400    Max.   :111968.4    Max.   :63.831    Max.   :84.774
##
##    rural.pop.perc        pop.tot        mobilesub_per100peeps
##    Min.   : 2.081    Min.   :2.548e+05    Min.   :   0.000
##    1st Qu.:21.992    1st Qu.:4.580e+06    1st Qu.:   1.437
##    Median :34.191    Median :1.042e+07    Median : 34.993
##    Mean   :36.904    Mean   :5.424e+07    Mean   : 52.295
##    3rd Qu.:49.272    3rd Qu.:3.854e+07    3rd Qu.:101.254
##    Max.   :87.379    Max.   :1.379e+09    Max.   :212.639
##
##    intl_tourist_arrival total_life_exp  life_expectancy_fe life_exp_male
##    Min.   :    12000    Min.   :43.41    Min.   :45.61    Min.   :41.38
##    1st Qu.:   739000    1st Qu.:68.97    1st Qu.:72.45    1st Qu.:65.47
##    Median :  2152000    Median :73.21    Median :76.37    Median :70.31
##    Mean   :  7075404    Mean   :71.67    Mean   :74.55    Mean   :68.90
##    3rd Qu.:  7045000    3rd Qu.:77.17    3rd Qu.:80.19    3rd Qu.:74.50
##    Max.   : 84452000    Max.   :83.98    Max.   :87.14    Max.   :81.70
##
##     trade_perGDP
##    Min.   : 13.75
##    1st Qu.: 51.67
##    Median : 70.85
##    Mean   : 80.51
##    3rd Qu.: 98.19
##    Max.   :408.36
##
```

```
# good. The data is cleaned. We can proceed to the next stage of the analysis.
```

```
## DATA EXPLORATORY ANALYSIS:
head(mergedata)
```

```
##       Country Year Max_Partners GDP_per_unit_CO2 PPP_Conv_Rate PPP_Share_GDP
## 1 Albania 1990           75          2.504851         2.117         0.035
## 2 Albania 1991           75          2.684573         2.775         0.024
## 3 Albania 1992           75          4.443426         9.488         0.020
## 4 Albania 1993           75          5.264840        19.912         0.022
## 5 Albania 1994           75          5.542105        26.714         0.023
## 6 Albania 1995           75          6.905429        28.740         0.024
##     Imports_PC Exports_PC Govt_Revenue gdp_per_cap agri_perc_gdp
## 1           0          0       -6.424    1838.673       36.4107
## 2           0          0       -6.424    1331.809       36.4107
## 3           0          0       -6.424    1243.609       36.4107
## 4           0          0       -6.424    1370.830       36.4107
## 5           0          0       -6.424    1493.790       36.4107
## 6           0          0       -6.424    1703.287       36.4107
##   agg.empl.agri.perc rural.pop.perc pop.tot mobilesub_per100peeps
## 1             55.914         63.572 3286542                     0
## 2             55.914         63.300 3266790                     0
## 3             56.134         62.751 3247039                     0
## 4             55.470         62.201 3227287                     0
## 5             54.841         61.646 3207536                     0
## 6             54.257         61.089 3187784                     0
##   intl_tourist_arrival total_life_exp life_expectancy_fe life_exp_male
## 1              1062000         71.836             74.991        69.070
## 2              1062000         71.803             74.980        69.017
## 3              1062000         71.802             74.985        68.997
## 4              1062000         71.860             75.039        69.037
## 5              1062000         71.992             75.158        69.150
## 6              1062000         72.205             75.352        69.347
##   trade_perGDP
## 1     39.43696
## 2     36.07052
## 3    108.78547
## 4     80.51833
## 5     53.10258
## 6     47.61059
```

```
summary(mergedata$Max_Partners)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    45.0   136.0   172.0   168.3   205.0   235.0
```

```r
# find top 10, middle 10, and bottom 10 with respect to the trading partners (Max_Par
tners var) in 2016
####################################
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────── tidyverse 1.3.0 ──
```

```
## ✓ ggplot2 3.3.0      ✓ purrr   0.3.4
## ✓ tibble  3.0.0      ✓ stringr 1.4.0
## ✓ tidyr   1.0.2      ✓ forcats 0.5.0
## ✓ readr   1.3.1
```

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

```r
library(ggplot2)
####################################

mergedata_2016 <- subset(mergedata, Year == 2016, select = c(Country,Max_Partners, Ye
ar))
mergedata_1990 <- subset(mergedata, Year == 1990, select = c(Country,Max_Partners, Ye
ar))

# top n by country for Max_Partners
top10_2016 <- mergedata %>%
  filter(Year == 2016) %>%
  top_n(n = 10, wt = Max_Partners)
```

```r
top10_1990 <- mergedata %>%
  filter(Year == 1990) %>%
  top_n(n = 10, wt = Max_Partners)


top10_1997 <- mergedata %>%
  filter(Year == 1997) %>%
  top_n(n = 10, wt = Max_Partners)


top10_2005 <- mergedata %>%
  filter(Year == 2005) %>%
  top_n(n = 10, wt = Max_Partners)


top10_2009 <- mergedata %>%
  filter(Year == 2009) %>%
  top_n(n = 10, wt = Max_Partners)


top10_2012 <- mergedata %>%
  filter(Year == 2012) %>%
  top_n(n = 10, wt = Max_Partners)

g90<- ggplot(data = top10_1990, mapping = aes(x = reorder(Country, Max_Partners), Max
_Partners)) +
  geom_bar(stat = "identity", width = 0.3) + coord_flip() +xlab("Country") + ylab("Nu
mber of Trading Partners")+ggtitle("Top 10 Countries in 1990")



g97<- ggplot(data = top10_1997, mapping = aes(x = reorder(Country, Max_Partners), Max
_Partners)) +
  geom_bar(stat = "identity", width = 0.3) + coord_flip() +xlab("Country") + ylab("Nu
mber of Trading Partners")+ggtitle("Top 10 Countries in 1997")

g05<- ggplot(data = top10_2005, mapping = aes(x = reorder(Country, Max_Partners), Max
_Partners)) +
  geom_bar(stat = "identity", width = 0.3) + coord_flip() +xlab("Country") + ylab("Nu
mber of Trading Partners")+ggtitle("Top 10 Countries in 2005")

g09<- ggplot(data = top10_2009, mapping = aes(x = reorder(Country, Max_Partners), Max
_Partners)) +
  geom_bar(stat = "identity", width = 0.3) + coord_flip() +xlab("Country") + ylab("Nu
mber of Trading Partners")+ggtitle("Top 10 Countries in 2009")

g12 <- ggplot(data = top10_2012, mapping = aes(x = reorder(Country, Max_Partners), Ma
x_Partners)) +
  geom_bar(stat = "identity", width = 0.3) + coord_flip() +xlab("Country") + ylab("Nu
mber of Trading Partners")+ggtitle("Top 10 Countries in 2012")
```

```
g16<- ggplot(data = top10_2016, mapping = aes(x = reorder(Country, Max_Partners), Max
_Partners)) +
  geom_bar(stat = "identity", width = 0.3) + coord_flip() +xlab("Country") + ylab("Nu
mber of Trading Partners")+ggtitle("Top 10 Countries in 2016")



#####################
library(ggpubr)
```
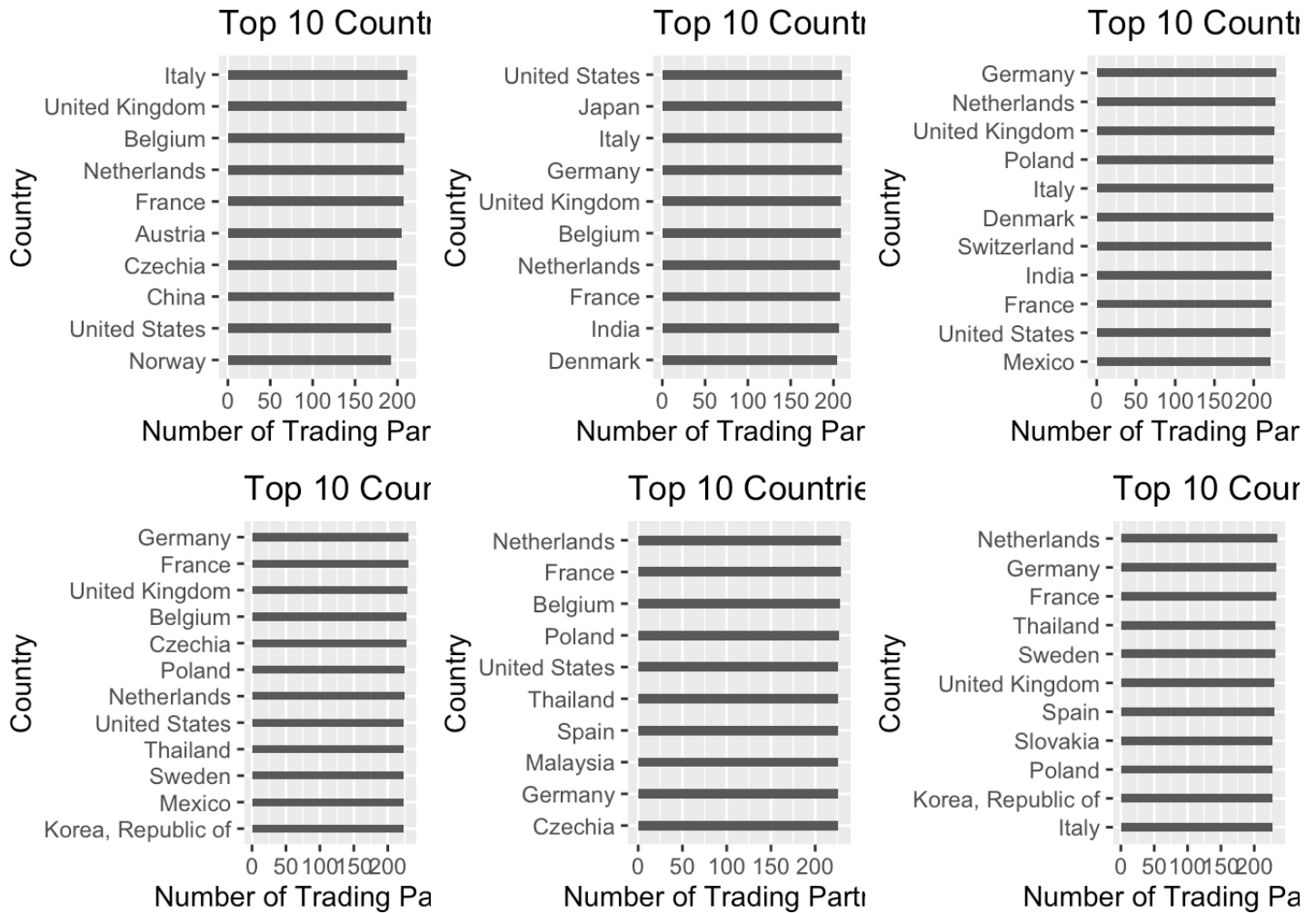
```
## Loading required package: magrittr
```

```
##
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':
##
##     set_names
```

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
####################

ggarrange(g90,g97,g05,g09,g12,g16)
```

Top 10 Counti

Top 10 Counti

Top 10 Counti

Top 10 Cour

Top 10 Countrie

Top 10 Cour


```
####################

bottom10_2016 <- mergedata %>%
  filter(Year == 2016) %>%
  top_n(n = -10, wt = Max_Partners) #top_n with neg value can be used to find the bottom values

bottom10_1990 <- mergedata %>%
  filter(Year == 1990) %>%
  top_n(n = -10, wt = Max_Partners)

bottom10_1997 <- mergedata %>%
  filter(Year == 1997) %>%
  top_n(n = -10, wt = Max_Partners)

bottom10_2005 <- mergedata %>%
  filter(Year == 2005) %>%
  top_n(n = -10, wt = Max_Partners)
```

```
bottom10_2009 <- mergedata %>%
  filter(Year == 2009) %>%
  top_n(n = -10, wt = Max_Partners)

bottom10_2012 <- mergedata %>%
  filter(Year == 2012) %>%
  top_n(n = -10, wt = Max_Partners)

b90<- ggplot(data = bottom10_1990, mapping = aes(x = reorder(Country, Max_Partners),
Max_Partners)) +
  geom_bar(stat = "identity", width = 0.3) + coord_flip() +xlab("Country") + ylab("Nu
mber of Trading Partners")+ggtitle("Lowest 10 Countries in 1990")


b97<- ggplot(data = bottom10_1997, mapping = aes(x = reorder(Country, Max_Partners),
Max_Partners)) +
  geom_bar(stat = "identity", width = 0.3) + coord_flip() +xlab("Country") + ylab("Nu
mber of Trading Partners")+ggtitle("Lowest 10 Countries in 1997")

b05<- ggplot(data = bottom10_2005, mapping = aes(x = reorder(Country, Max_Partners),
Max_Partners)) +
  geom_bar(stat = "identity", width = 0.3) + coord_flip() +xlab("Country") + ylab("Nu
mber of Trading Partners")+ggtitle("Lowest 10 Countries in 2005")

b09<- ggplot(data = bottom10_2009, mapping = aes(x = reorder(Country, Max_Partners),
Max_Partners)) +
  geom_bar(stat = "identity", width = 0.3) + coord_flip() +xlab("Country") + ylab("Nu
mber of Trading Partners")+ggtitle("Lowest 10 Countries in 2009")

b12 <- ggplot(data = bottom10_2012, mapping = aes(x = reorder(Country, Max_Partners),
Max_Partners)) +
  geom_bar(stat = "identity", width = 0.3) + coord_flip() +xlab("Country") + ylab("Nu
mber of Trading Partners")+ggtitle("Lowest 10 Countries in 2012")

b16<- ggplot(data = bottom10_2016, mapping = aes(x = reorder(Country, Max_Partners),
Max_Partners)) +
  geom_bar(stat = "identity", width = 0.3) + coord_flip() +xlab("Country") + ylab("Nu
mber of Trading Partners")+ggtitle("Lowest 10 Countries in 2016")

#######
ggarrange(b90,b97,b05,b09,b12,b16)
```
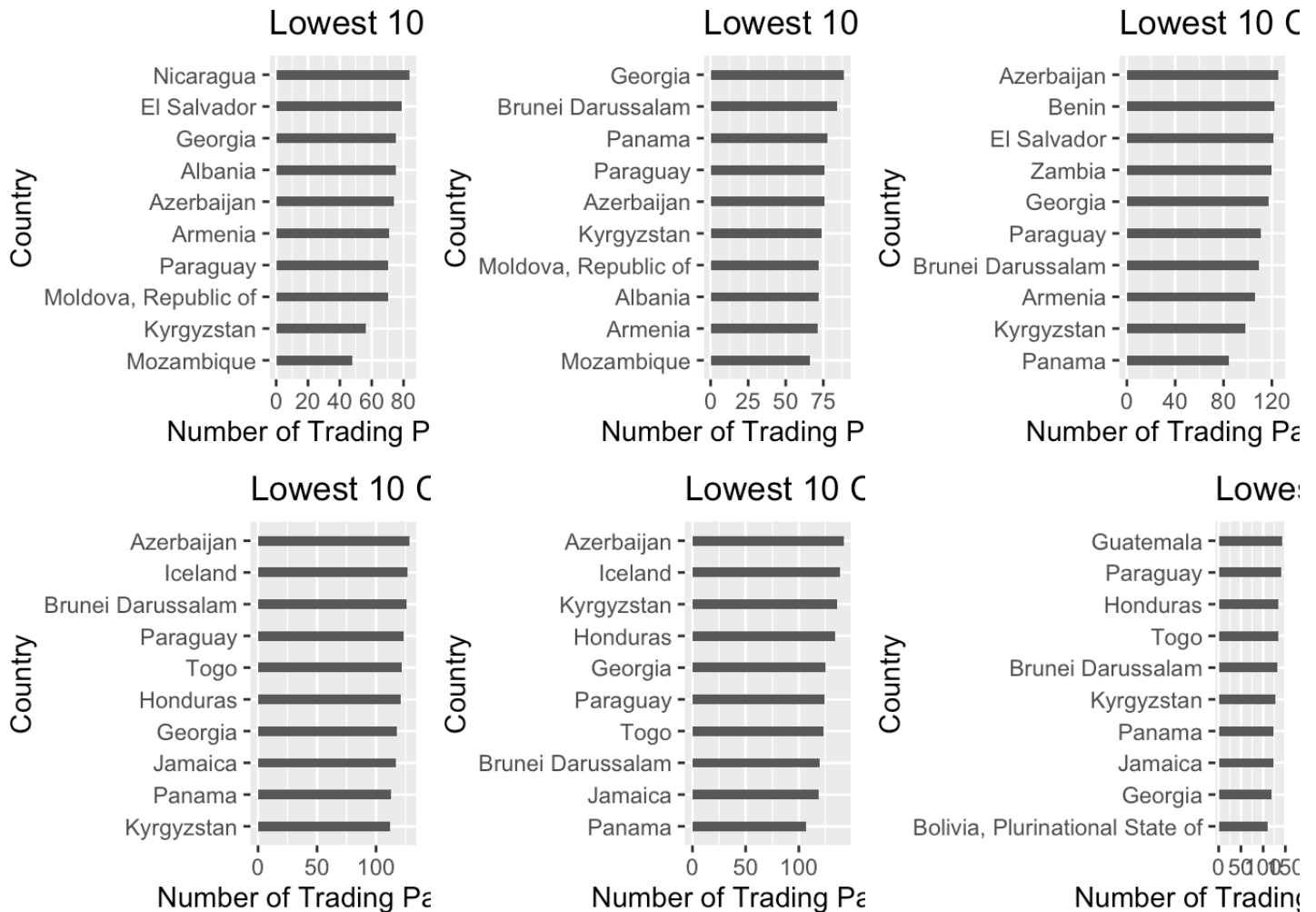
Lowest 10

Country	
Nicaragua	
El Salvador	
Georgia	
Albania	
Azerbaijan	
Armenia	
Paraguay	
Moldova, Republic of	
Kyrgyzstan	
Mozambique	

Number of Trading P (0 20 40 60 80)

Lowest 10

Country	
Georgia	
Brunei Darussalam	
Panama	
Paraguay	
Azerbaijan	
Kyrgyzstan	
Moldova, Republic of	
Albania	
Armenia	
Mozambique	

Number of Trading P (0 25 50 75)

Lowest 10 C

Country	
Azerbaijan	
Benin	
El Salvador	
Zambia	
Georgia	
Paraguay	
Brunei Darussalam	
Armenia	
Kyrgyzstan	
Panama	

Number of Trading Pa (0 40 80 120)

Lowest 10 C

Country	
Azerbaijan	
Iceland	
Brunei Darussalam	
Paraguay	
Togo	
Honduras	
Georgia	
Jamaica	
Panama	
Kyrgyzstan	

Number of Trading Pa (0 50 100)

Lowest 10 C

Country	
Azerbaijan	
Iceland	
Kyrgyzstan	
Honduras	
Georgia	
Paraguay	
Togo	
Brunei Darussalam	
Jamaica	
Panama	

Number of Trading Pa (0 50 100)

Lowes

Country	
Guatemala	
Paraguay	
Honduras	
Togo	
Brunei Darussalam	
Kyrgyzstan	
Panama	
Jamaica	
Georgia	
Bolivia, Plurinational State of	

Number of Trading (0 50 100 150)

```
######
# Average across the time span:
avg_partners <- mergedata %>% group_by(Country) %>% summarise_at(vars(Max_Partners),f
uns(mean(.,na.rm=TRUE)))
```

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## Please use a list of either functions or lambdas:
##
##    # Simple named list:
##    list(mean = mean, median = median)
##
##    # Auto named with `tibble::lst()`:
##    tibble::lst(mean, median)
##
##    # Using lambdas
##    list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once per session.
```

```
avg_top10 <- avg_partners %>%
  top_n(n = 10, wt = Max_Partners)

avg_bottom10 <- avg_partners %>% top_n(n = -10, wt = Max_Partners)

avg10_top<- ggplot(data = avg_top10, mapping = aes(x = reorder(Country, Max_Partners)
, Max_Partners)) +
  geom_bar(stat = "identity", width = 0.3) + coord_flip() +xlab("Country") + ylab("Av
g Number of Trading Partners")+ggtitle("Top 10 Countries: 1990-2016")

avg10_bottom<- ggplot(data = avg_bottom10, mapping = aes(x = reorder(Country, Max_Par
tners), Max_Partners)) +
  geom_bar(stat = "identity", width = 0.3) + coord_flip() +xlab("Country") + ylab("Av
g Number of Trading Partners")+ggtitle("Lowest 10 Countries: 1990-2016")

ggarrange(avg10_bottom,avg10_top)
```
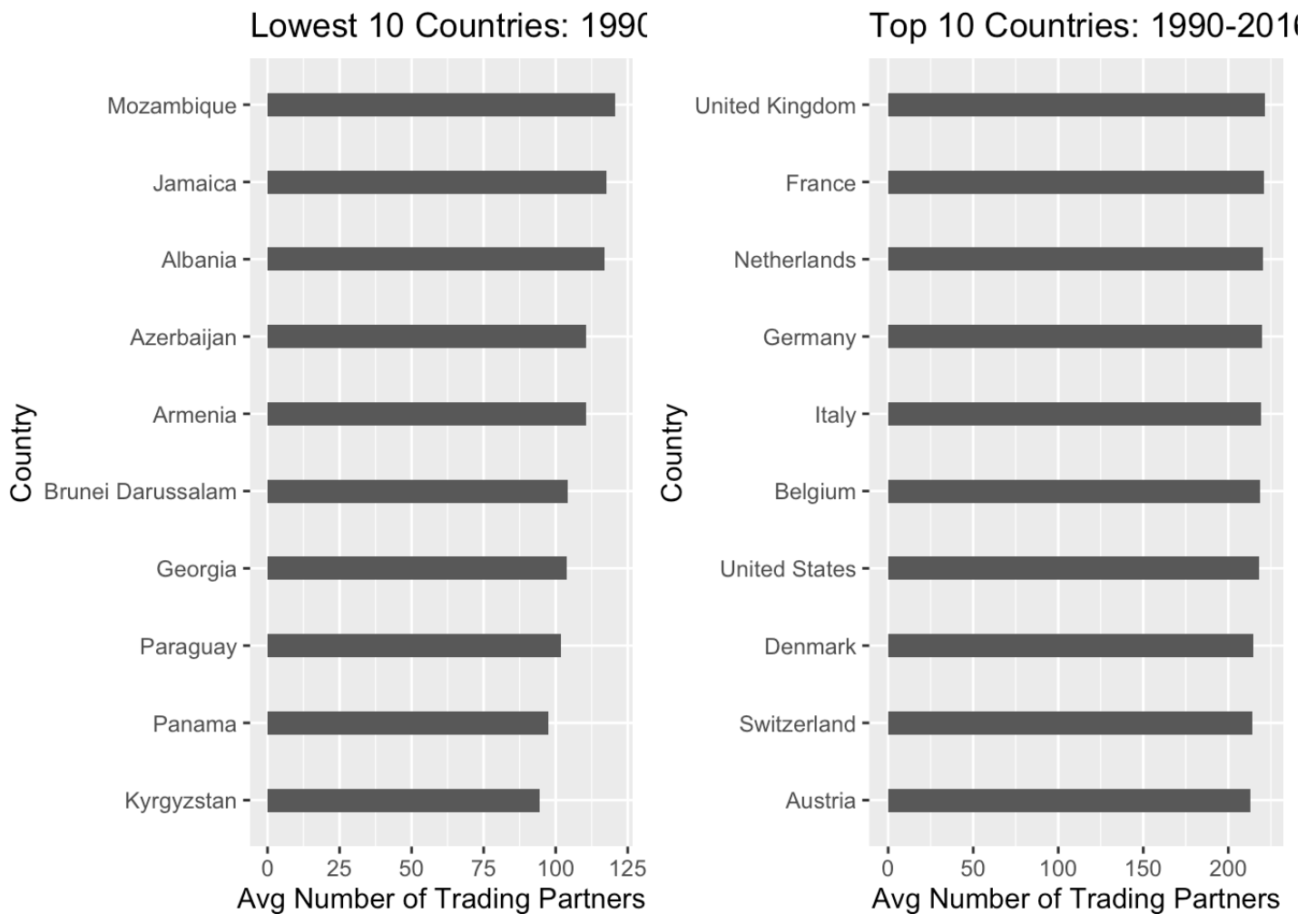
```
#############################
## Cases of Thailand and South Korea:

data_thailand <- subset(mergedata, Country == "Thailand")
data_southkorea <- subset(mergedata, Country == "Korea, Republic of")
head(data_thailand)
```

```
##         Country Year Max_Partners GDP_per_unit_CO2 PPP_Conv_Rate PPP_Share_GDP
## 2377 Thailand 1990          163         4.557155         9.391         0.880
## 2378 Thailand 1991          169         4.445433         9.564         0.932
## 2379 Thailand 1992          174         4.405588         9.728         0.914
## 2380 Thailand 1993          192         4.231155         9.721         0.975
## 2381 Thailand 1994          195         4.093368         9.962         1.022
## 2382 Thailand 1995          201         3.906547        10.318         1.066
##      Imports_PC Exports_PC Govt_Revenue gdp_per_cap agri_perc_gdp
## 2377     23.682     11.686       -0.777    2503.803     12.499628
## 2378     12.950     17.280       -0.777    2686.062     12.649827
## 2379      8.968     13.807       -0.777    2874.132     12.297336
## 2380     11.778     12.735       -0.777    3083.186      8.027850
## 2381     15.749     14.245       -0.777    3299.346      8.629039
## 2382     19.968     15.445       -0.777    3531.749      9.082753
##      agg.empl.agri.perc rural.pop.perc  pop.tot mobilesub_per100peeps
## 2377             60.334         70.576 56558186             0.1117840
## 2378             60.334         70.407 57232465             0.2158757
## 2379             60.905         70.237 57811021             0.4334537
## 2380             56.773         70.066 58337773             0.7089009
## 2381             55.988         69.895 58875269             1.2522796
## 2382             51.975         69.724 59467274             2.1824205
##      intl_tourist_arrival total_life_exp life_expectancy_fe life_exp_male
## 2377              6952000         70.248             73.433        67.175
## 2378              6952000         70.300             73.608        67.117
## 2379              6952000         70.281             73.727        66.976
## 2380              6952000         70.239             73.823        66.815
## 2381              6952000         70.202             73.915        66.668
## 2382              6952000         70.191             74.013        66.565
##      trade_perGDP
## 2377     75.78236
## 2378     78.47113
## 2379     77.95465
## 2380     77.74581
## 2381     81.24895
## 2382     89.75628
```

```
old.par <- par(mfrow=c(2, 3))
plot(data_thailand$gdp_per_cap,data_thailand$Max_Partners, col = "red",xlab = "GDP Pe
r Capita (US Dollars)",ylab = "Number of Trading Partners", main = "Case 1: Thailand"
)
plot(data_thailand$GDP_per_unit_CO2,data_thailand$Max_Partners, col = "red",xlab = "G
DP per CO2 unit",ylab = "Number of Trading Partners", main = "Case 1: Thailand")
plot(data_thailand$agri_perc_gdp,data_thailand$Max_Partners, col = "red",xlab = "Agri
cultural Percentage of GDP",ylab = "Number of Trading Partners", main = "Case 1: Thai
land")
plot(data_thailand$total_life_exp,data_thailand$Max_Partners, col = "red",xlab = "Ave
rage Life Expectancy",ylab = "Number of Trading Partners", main = "Case 1: Thailand")
plot(data_thailand$intl_tourist_arrival,data_thailand$Max_Partners, col = "red",xlab
= "Total Intl Tourist Arrival",ylab = "Number of Trading Partners", main = "Case 1: T
hailand")
plot(data_thailand$mobilesub_per100peeps,data_thailand$Max_Partners, col = "red",xlab
= "Mobile Subscription per 100 People",ylab = "Number of Trading Partners", main = "C
ase 1: Thailand")
```
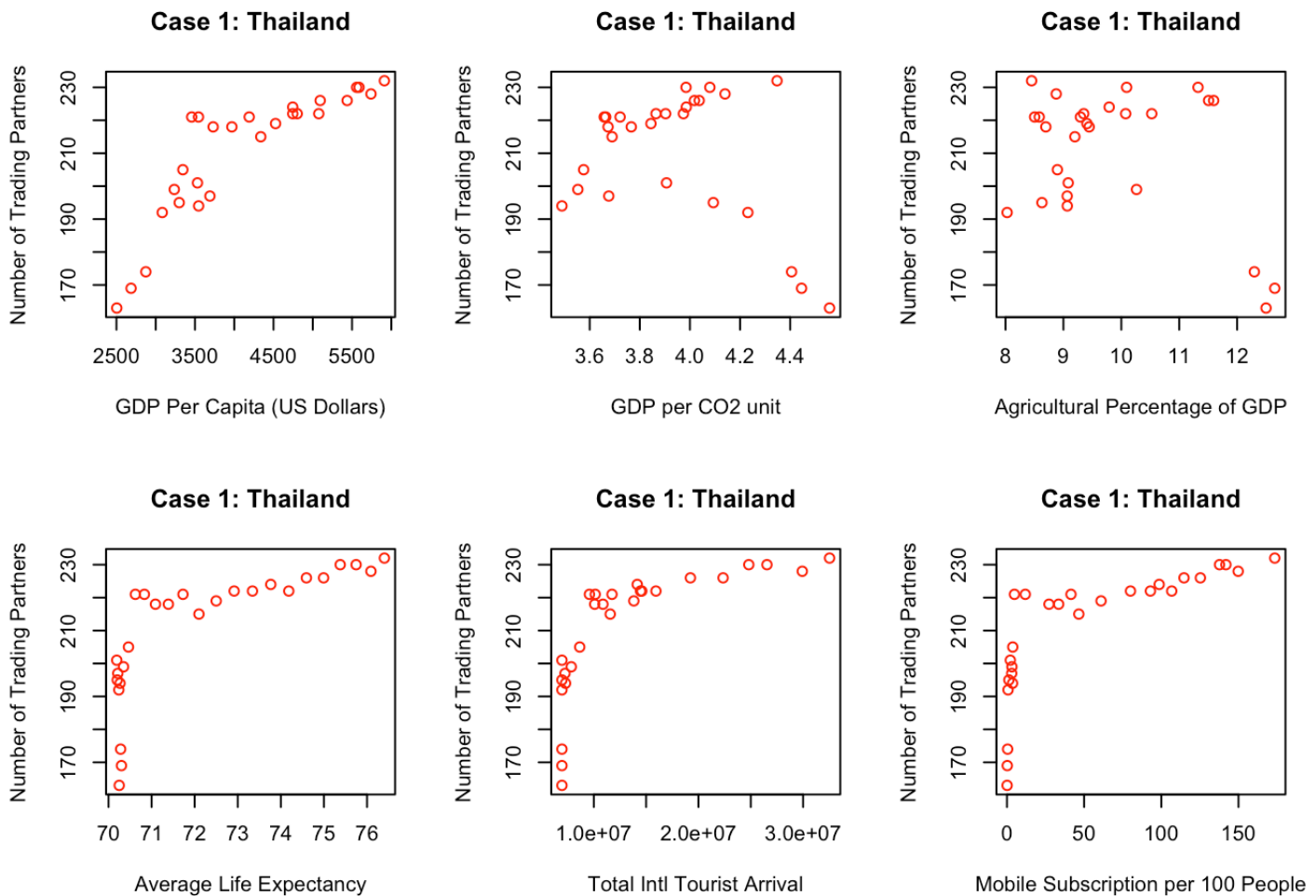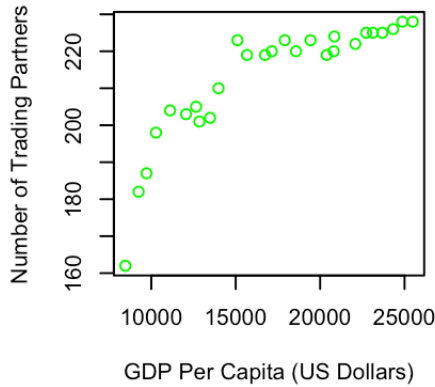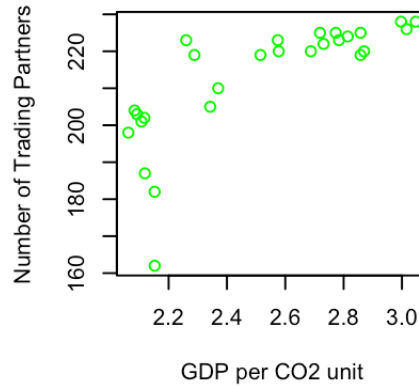
```
par(old.par)

old.par2 <- par(mfrow=c(2, 3))
plot(data_southkorea$gdp_per_cap,data_southkorea$Max_Partners, col = "green",xlab = "
GDP Per Capita (US Dollars)", ylab = "Number of Trading Partners", main = "Case 2: So
uth Korea")
plot(data_southkorea$GDP_per_unit_CO2,data_southkorea$Max_Partners, col = "green",xla
b = "GDP per CO2 unit", ylab = "Number of Trading Partners", main = "Case 2: South Ko
rea")
plot(data_southkorea$agri_perc_gdp,data_southkorea$Max_Partners, col = "green",xlab =
"Agricultural Percentage of GDP", ylab = "Number of Trading Partners", main = "Case 2
: South Korea")
plot(data_southkorea$total_life_exp,data_southkorea$Max_Partners, col = "green",xlab
= "Average Life Expectancy", ylab = "Number of Trading Partners", main = "Case 2: Sou
th Korea")
plot(data_southkorea$intl_tourist_arrival,data_southkorea$Max_Partners, col = "green"
,xlab = "Total Intl Tourist Arrival", ylab = "Number of Trading Partners", main = "Ca
se 2: South Korea")
plot(data_southkorea$mobilesub_per100peeps,data_southkorea$Max_Partners, col = "green
",xlab = "Mobile Subscription per 100 People", ylab = "Number of Trading Partners", m
ain = "Case 2: South Korea")
```
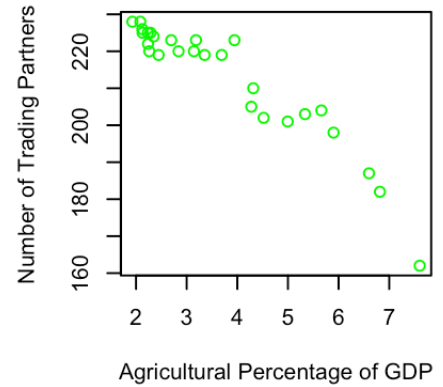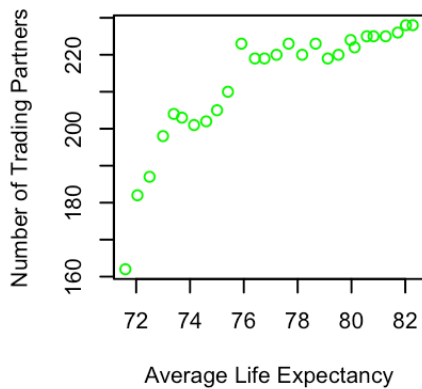
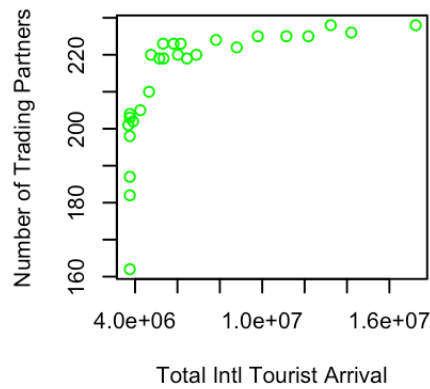Case 2: South Korea — Number of Trading Partners vs. GDP Per Capita (US Dollars)

Case 2: South Korea — Number of Trading Partners vs. GDP per CO2 unit

Case 2: South Korea — Number of Trading Partners vs. Agricultural Percentage of GDP

Case 2: South Korea — Number of Trading Partners vs. Average Life Expectancy

Case 2: South Korea — Number of Trading Partners vs. Total Intl Tourist Arrival

Case 2: South Korea — Number of Trading Partners vs. Mobile Subscription per 100 People

```
par(old.par2)

############ Principal Component Analysis:

## First, we need to use the stationary data over the entire time span to do PCA inst
ead of using time-series data.

avg_all <- mergedata %>% group_by(Country) %>% summarise_at(vars(Max_Partners:trade_p
erGDP), mean, na.rm=TRUE)
avg_all <- data.frame(avg_all)

avg_all$ID <- seq.int(nrow(avg_all))
country_id <- subset(avg_all, select=c("Country","ID"))

sapply(avg_all, class) # check to see if numeric before PCA.
```

```
##               Country        Max_Partners       GDP_per_unit_CO2
##              "factor"           "numeric"              "numeric"
##         PPP_Conv_Rate        PPP_Share_GDP             Imports_PC
##             "numeric"           "numeric"              "numeric"
##            Exports_PC         Govt_Revenue            gdp_per_cap
##             "numeric"           "numeric"              "numeric"
##         agri_perc_gdp     agg.empl.agri.perc        rural.pop.perc
##             "numeric"           "numeric"              "numeric"
##               pop.tot mobilesub_per100peeps   intl_tourist_arrival
##             "numeric"           "numeric"              "numeric"
##        total_life_exp     life_expectancy_fe           life_exp_male
##             "numeric"           "numeric"              "numeric"
##          trade_perGDP                   ID
##             "numeric"           "integer"
```

```r
# remove "country"

avg_all_no_countryname <- subset(avg_all, select=c(-Country))

# turning the ID column into row name instead of a column
samp2 <- avg_all_no_countryname[,-19]
rownames(samp2) <- avg_all_no_countryname[,19]

pca_data <- subset(samp2, select = c(-Max_Partners))
pr.out = prcomp(pca_data, scale = TRUE)

names(pr.out)
```

```
## [1] "sdev"     "rotation" "center"    "scale"     "x"
```

```r
pr.out$center
```

```
##         GDP_per_unit_CO2           PPP_Conv_Rate            PPP_Share_GDP
##               5.134184e+00            8.619720e+01             9.432858e-01
##                  Imports_PC              Exports_PC             Govt_Revenue
##               6.591844e+00            5.791484e+00            -1.627273e-02
##                 gdp_per_cap            agri_perc_gdp          agg.empl.agri.perc
##               1.605263e+04            9.791403e+00             2.240515e+01
##               rural.pop.perc                 pop.tot      mobilesub_per100peeps
##               3.690374e+01            5.423977e+07             5.229527e+01
##        intl_tourist_arrival          total_life_exp          life_expectancy_fe
##               7.075404e+06            7.166571e+01             7.454690e+01
##                life_exp_male             trade_perGDP
##               6.890341e+01            8.050765e+01
```

pr.out$scale

```
##         GDP_per_unit_CO2           PPP_Conv_Rate            PPP_Share_GDP
##               3.174002e+00            2.838482e+02             2.324133e+00
##                  Imports_PC              Exports_PC             Govt_Revenue
##               3.080442e+00            3.259744e+00             2.121656e+00
##                 gdp_per_cap            agri_perc_gdp          agg.empl.agri.perc
##               1.920443e+04            8.917634e+00             1.965072e+01
##               rural.pop.perc                 pop.tot      mobilesub_per100peeps
##               1.861833e+01            1.714795e+08             1.765922e+01
##        intl_tourist_arrival          total_life_exp          life_expectancy_fe
##               1.199943e+07            7.625178e+00             7.997949e+00
##                life_exp_male             trade_perGDP
##               7.447609e+00            4.049715e+01
```

pr.out$rotation

```
##                               PC1          PC2          PC3          PC4
## GDP_per_unit_CO2       -0.22104184   0.13088713  -0.17977715  -0.306270870
## PPP_Conv_Rate          -0.06370749   0.01415015  -0.09116653  -0.474835678
## PPP_Share_GDP           0.07006169  -0.55388903  -0.18870033  -0.046481800
## Imports_PC             -0.09328565  -0.18135920   0.60179388  -0.248026788
## Exports_PC             -0.06559144  -0.21658963   0.62694760  -0.155949832
## Govt_Revenue            0.04641659   0.17303334  -0.05867869  -0.676410820
## gdp_per_cap             0.27649157   0.01862988  -0.07747907  -0.135168722
## agri_perc_gdp          -0.33888797  -0.00221355  -0.01705228   0.131116455
## agg.empl.agri.perc     -0.35342167  -0.04430116  -0.01405742   0.021551453
## rural.pop.perc         -0.31407920  -0.05941447   0.02145071   0.144993974
## pop.tot                -0.05774140  -0.51936282   0.00938868   0.017381679
## mobilesub_per100peeps   0.33293425   0.07253167   0.10428850  -0.062307144
## intl_tourist_arrival    0.13397772  -0.44417899  -0.19130813  -0.029527203
```

```
## total_life_exp            0.35218987 -0.03647493  0.03806898  0.033488717
## life_expectancy_fe        0.35156580 -0.03646336  0.06342725  0.061852092
## life_exp_male             0.34581696 -0.03617717  0.01197747  0.002586423
## trade_perGDP              0.11319941  0.28987156  0.32398824  0.261227086
##                                  PC5          PC6          PC7          PC8
## GDP_per_unit_CO2         -0.128415266 -0.62186064 -0.20358913  0.076046905
## PPP_Conv_Rate             0.776380135 -0.10200189  0.21816519 -0.271105455
## PPP_Share_GDP            -0.070587417 -0.03041090  0.16954950 -0.186944672
## Imports_PC               -0.070307493  0.07222033 -0.21467075 -0.185012484
## Exports_PC               -0.006238384 -0.21855785 -0.07953351  0.090086386
## Govt_Revenue             -0.346330480  0.31761176  0.27151982  0.310691284
## gdp_per_cap              -0.270127073 -0.48088140  0.20003499 -0.005394304
## agri_perc_gdp             0.061357606 -0.16070319 -0.02009671  0.140027346
## agg.empl.agri.perc        0.012031810 -0.11649623  0.05127892  0.226271478
## rural.pop.perc            0.001110643 -0.20460380  0.15363706  0.270695498
## pop.tot                   0.059907503  0.11043591  0.41619067  0.386374641
## mobilesub_per100peeps    -0.122720202 -0.07172726  0.01237027 -0.111376141
## intl_tourist_arrival     -0.193281749 -0.18517231 -0.05817219 -0.330131119
## total_life_exp            0.200399102 -0.10657608 -0.12276407  0.312960547
## life_expectancy_fe        0.195842879 -0.07051640 -0.14676484  0.253089174
## life_exp_male             0.197523878 -0.14177652 -0.09094713  0.366310739
## trade_perGDP             -0.020436815 -0.21801788  0.68395212 -0.189541539
##                                  PC9         PC10         PC11         PC12
## GDP_per_unit_CO2          0.17017139 -0.489709094  0.04829099 -0.218648030
## PPP_Conv_Rate            -0.04466353  0.112854574  0.09250038  0.053684426
## PPP_Share_GDP             0.11702435  0.060504294 -0.47552721 -0.455530556
## Imports_PC                0.16403288 -0.284266830  0.01633140  0.295167561
## Exports_PC               -0.14850583  0.325479039 -0.10559283 -0.282774283
## Govt_Revenue             -0.28537225 -0.008281377 -0.13904031  0.016854766
## gdp_per_cap               0.34170382  0.432413526 -0.02513892  0.326282514
## agri_perc_gdp            -0.06143913  0.227137034 -0.27854181  0.439391286
## agg.empl.agri.perc       -0.13530008  0.155906559 -0.11344498  0.146148265
## rural.pop.perc           -0.35472249  0.114413780  0.37377244 -0.283674968
## pop.tot                   0.35216613 -0.230242058  0.34540565  0.164394460
## mobilesub_per100peeps     0.01916796  0.256630356  0.53252175 -0.156045993
## intl_tourist_arrival     -0.61067900 -0.156170935  0.15069043  0.337509412
## total_life_exp           -0.11155224 -0.097996342 -0.11274389  0.032060543
## life_expectancy_fe       -0.14331176 -0.082011771 -0.06245623 -0.017569002
## life_exp_male            -0.07536078 -0.112042991 -0.15595122  0.080309200
## trade_perGDP             -0.15590070 -0.329943105 -0.18686580 -0.001808859
##                                 PC13         PC14          PC15          PC16
## GDP_per_unit_CO2          0.079122659 -0.171149543  0.0286076504  0.046596989
## PPP_Conv_Rate            -0.024169918  0.028721645  0.0003148129  0.010842994
## PPP_Share_GDP            -0.330323975 -0.131710519 -0.0526051193 -0.026619776
## Imports_PC               -0.473778501  0.116091957 -0.1005260366 -0.033278741
## Exports_PC                0.477891216 -0.025764460  0.1443274421  0.013391477
## Govt_Revenue             -0.061253429 -0.057980098  0.0705035245  0.051472000
```

```
## gdp_per_cap            -0.064162829  0.361567158 -0.0165068093  0.093368755
## agri_perc_gdp          -0.215394582 -0.517837612  0.4148087576  0.073707243
## agg.empl.agri.perc      0.034310381 -0.098453290 -0.8410645065 -0.099031712
## rural.pop.perc         -0.474590643  0.354077221  0.1669867724 -0.001944363
## pop.tot                 0.210614995 -0.111434525  0.0635515970  0.063828136
## mobilesub_per100peeps  -0.219199401 -0.612250463 -0.1437537302 -0.136551022
## intl_tourist_arrival    0.169681021  0.009395165  0.0152259276 -0.014726346
## total_life_exp         -0.096098751 -0.007559637 -0.0416536151  0.033420746
## life_expectancy_fe     -0.126253407 -0.056882726 -0.1390599131  0.697872189
## life_exp_male          -0.064352189  0.044090009  0.0720007550 -0.676774029
## trade_perGDP           -0.001857863 -0.092134651 -0.0412569453 -0.001455884
##                                  PC17
## GDP_per_unit_CO2       -2.697424e-04
## PPP_Conv_Rate          -2.133766e-04
## PPP_Share_GDP           2.707183e-05
## Imports_PC              7.642497e-05
## Exports_PC              4.191007e-04
## Govt_Revenue            2.866759e-04
## gdp_per_cap            -1.683234e-04
## agri_perc_gdp          -8.356154e-05
## agg.empl.agri.perc     -3.863747e-03
## rural.pop.perc          2.288787e-03
## pop.tot                 8.564887e-04
## mobilesub_per100peeps   2.512656e-03
## intl_tourist_arrival    3.758977e-04
## total_life_exp          8.121559e-01
## life_expectancy_fe     -4.257813e-01
## life_exp_male          -3.988545e-01
## trade_perGDP            6.007918e-04
```

```
biplot(pr.out, scale = 0)
```
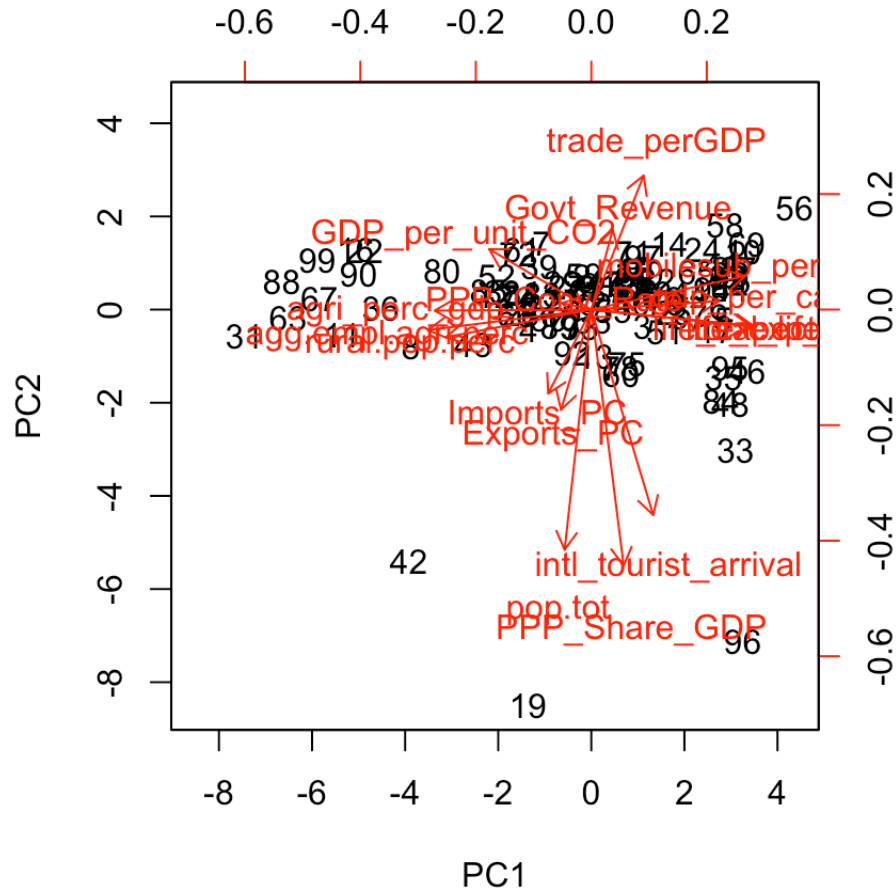
```
#### Instead of doing PCA for all variables, we are going to include some selected va
riables from the exploratory analysis section.

subset_pcadata <- subset(mergedata,select =c("Country", "Year", "gdp_per_cap","GDP_pe
r_unit_CO2","agri_perc_gdp","total_life_exp","intl_tourist_arrival","mobilesub_per100
peeps","agg.empl.agri.perc"))

subset_pcaavg <- subset_pcadata %>% group_by(Country) %>% summarise_at(vars(gdp_per_c
ap:agg.empl.agri.perc),mean,na.rm=TRUE)

subset_pcaavg <- data.frame(subset_pcaavg)

subset_pcaavg$ID <- seq.int(nrow(subset_pcaavg))
country_id2 <- subset(subset_pcaavg, select=c("Country","ID"))

sapply(subset_pcaavg, class) # check to see if numeric before PCA.
```

```
##                Country          gdp_per_cap      GDP_per_unit_CO2
##               "factor"           "numeric"            "numeric"
##          agri_perc_gdp       total_life_exp   intl_tourist_arrival
##              "numeric"           "numeric"            "numeric"
## mobilesub_per100peeps      agg.empl.agri.perc                    ID
##              "numeric"           "numeric"            "integer"
```

```
# remove "country"

subset_pca_no_countryname <- subset(subset_pcaavg, select=c(-Country))

# turning the ID column into row name instead of a column
samp3 <- subset_pca_no_countryname[,-8]
rownames(samp3) <- subset_pca_no_countryname[,8]

pr.out3 = prcomp(samp3, scale = TRUE)

pr.out3$rotation
```

```
##                                 PC1          PC2          PC3          PC4
## gdp_per_cap              0.3543596 -0.17954409 -0.55251248  0.66026780
## GDP_per_unit_CO2        -0.2885919 -0.27103080 -0.75656495 -0.49812578
## agri_perc_gdp           -0.4351674 -0.07567857  0.01551139  0.46716951
## total_life_exp           0.4241552  0.05188499  0.01631303 -0.10036049
## intl_tourist_arrival     0.1905211 -0.92531747  0.30548353 -0.06432859
## mobilesub_per100peeps    0.4311152  0.11166338 -0.16775681  0.02121868
## agg.empl.agri.perc      -0.4468860 -0.13117088  0.01924012  0.28811309
##                                 PC5          PC6          PC7
## gdp_per_cap             -0.06136437  0.31110275  0.02047082
## GDP_per_unit_CO2         0.11260265 -0.08888081 -0.04645471
## agri_perc_gdp            0.31554845 -0.41268703 -0.56261243
## total_life_exp           0.89020008 -0.04305645  0.11292441
## intl_tourist_arrival    -0.04880529 -0.07915067 -0.03720449
## mobilesub_per100peeps   -0.26254540 -0.82964801  0.12582723
## agg.empl.agri.perc       0.14218302 -0.16902717  0.80679711
```

```
biplot(pr.out3, scale = 0)
```