Visa Network Data

Lisa Reiber

4 11 2017

```
#Get Data
visa.df <- readRDS(file = "data/VisaNetworkData_041017")
#Load Packages
if (!require("pacman")) install.packages("pacman")
pacman::p_load(tidyverse)
```

Question 1

In the session, we already learned that mobility rights (variable: indegree) are unequally distributed among continents (continent). We can check once more using the following code:

```
arrange(summarize(group_by(visa.df, continent), mean = mean(indegree), n()), desc(mean))
```

Answer 1:

```
visa.df %>%
      group_by(continent) %>%
      dplyr:: summarize(mean = mean(indegree), n()) %>%
      arrange(desc(mean)) %>%
      as.data.frame()
```

```
##   continent    mean n()
## 1    Europe 63.4359  39
## 2  Americas 42.5200  25
## 3   Oceania 40.0000   5
## 4      Asia 19.7234  47
## 5    Africa 14.5600  50
```

Question 2

Recode the variable gdppc into deciles. Then group the data by this variable and compute the mean indegree. Are gdppc and indegree related? You can also check the correlation between the variables (?cor)

Hints: Have a look at the function ntile or cut (rather go with ntile; it is easier). Try to remove the missing values beforehand !is.na()

Answer 2:

```
# data cleaning and description
      visa.df %>%
            filter(!is.na(gdppc) & !is.na(indegree)) %>%
            mutate(dezgdp = ntile(gdppc, 10)) %>%
            group_by(dezgdp) %>%
            dplyr:: summarize(mean.indegree = mean(indegree),
                cor.gdp.indegree = cor(gdppc, indegree), n())
```

```
## # A tibble: 10 x 4
##    dezgdp mean.indegree cor.gdp.indegree `n()`
##     <int>         <dbl>            <dbl> <int>
##  1      1      13.12500        0.2597284    16
##  2      2      14.00000       -0.2144792    16
##  3      3      15.46667        0.1010284    15
##  4      4      18.87500        0.3152741    16
##  5      5      12.26667       -0.1186521    15
##  6      6      31.75000        0.3964359    16
##  7      7      45.00000        0.3264340    16
##  8      8      61.33333        0.1277759    15
##  9      9      69.87500       -0.1970118    16
## 10     10      59.33333       -0.3487181    15
```

```
# Correlation of GDP per Capita and Number of Visa Waivers Granted
      visa.df %>%
            filter(!is.na(gdppc) & !is.na(indegree)) %>%
            dplyr::summarize(cor.gdp.indegree = cor(gdppc, indegree), n = n()) %>%
            as.data.frame()
```

```
##   cor.gdp.indegree   n
## 1        0.5984022 156
```

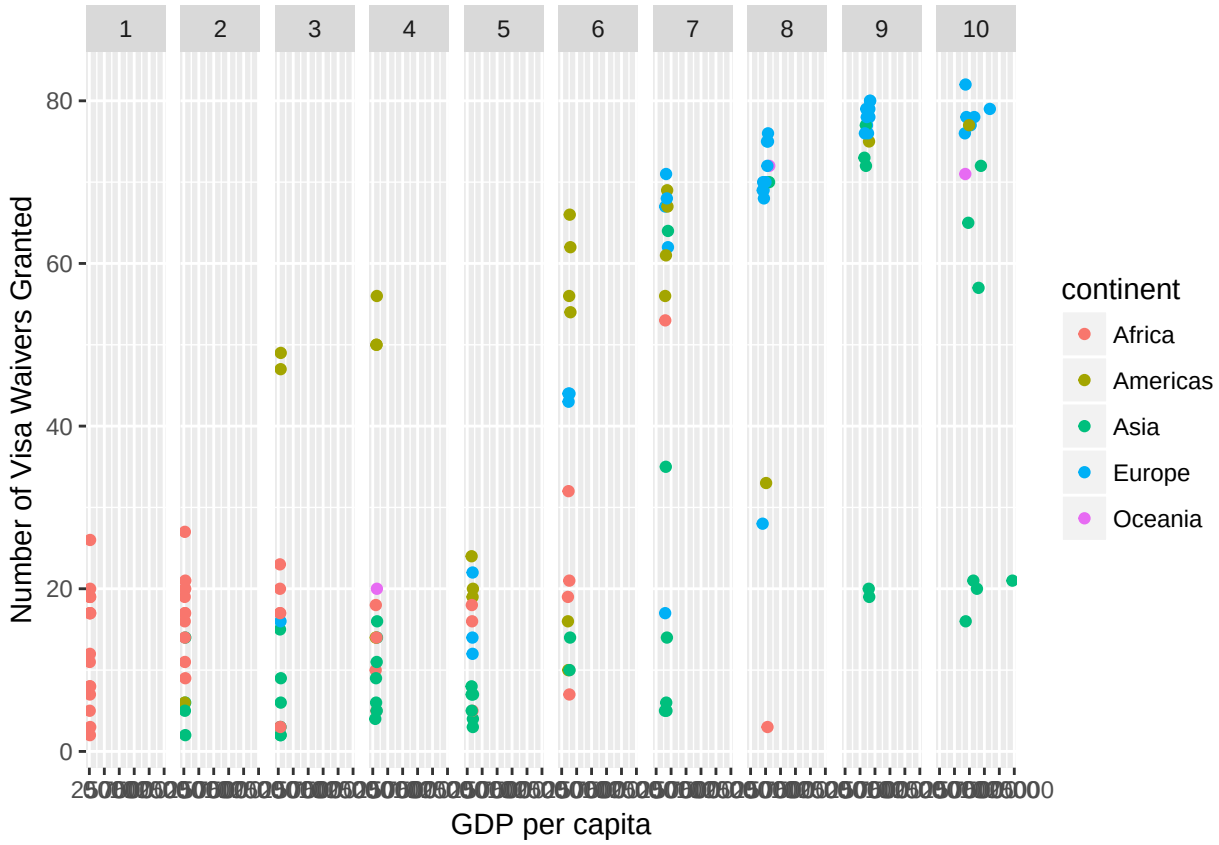
Extra: Plotting the relation

```
# scatter plot by deziles

visa.df <- visa.df %>%
            filter(!is.na(gdppc) & !is.na(indegree)) %>%
            mutate(dezgdp = ntile(gdppc, 10))

ggplot(visa.df, aes(x = gdppc ,
                y = indegree,
                color = continent)) +
      geom_point() +
```

```
      scale_y_continuous(name = "Number of Visa Waivers Granted") +
      scale_x_continuous(name = "GDP per capita") +
      facet_grid(~dezgdp)
```




```
# scatterplot overall
ggplot(visa.df, aes(x = gdppc ,
                    y = indegree,
                    color = continent)) +
      geom_point() +
      geom_text(data = subset(visa.df, gdppc > 50000),
                aes(gdppc, indegree, label=name)) +
      scale_y_continuous(name = "Number of Visa Waivers Granted") +
      scale_x_continuous(name = "GDP per capita") +
      geom_smooth(method=lm,
                  linetype="dashed")
```


Number of Visa Waivers Granted
200
100
0
−100
0
25000
50000
75000
100000
125000
GDP per capita
Luxembourg
Singapore
Brunei
Qatar
continent
Africa
Americas
Asia
Europe
Oceania