

Statistics on Illiteracy and Rape crime for US State

kjarnutowska

19 października 2017

In this exercise I am working with two datasets. Data sets related to the 50 states of the United States of America. Both are from **datasets** package. First **state.x77** is giving some basics statistics like **Population** (population estimate as of July 1, 1975), **Income** (per capita income, 1974), **Illiteracy** (1970, percent of population), **Life Exp** (life expectancy in years 1969-71), **Murder** (murder and non-negligent manslaughter rate per 100,000 population, 1976), **HS Grad** (percent high-school graduates, 1970), **Frost** (mean number of days with minimum temperature below freezing (1931-1960) in capital or large city) and **Area** (in square miles). Second is **USArrests**. This data set contains statistics (**Murder**, **Assault**, **UrbanPop**, **Rape**), in arrests per 100,000 in each of the 50 US states in 1973.

3 biggest states from choosen ones (with area limit)

```
library(datasets)
data(state)

library(dplyr)
```

Working on **state.x77** I am changing data set class from matrix to data frame.

```
state.x77 <- as.data.frame(state.x77)
state.x77$State <- rownames(state.x77)
summary(state.x77) # summary to chcek if there are NaN's
```

Now I want to filter states with area between 50000 and 200000. Next is to select names of tree the biggest ones.

```
stanyZPowierzchniaMiedzi50ka200k = state.x77 %>%
  filter(Area >=50000 & Area <=200000) %>%
  arrange(desc(Area))
stanyZPowierzchniaMiedzi50ka200k %>%
  select(State) %>%
  head(3)
```

```
##      State
## 1 California
## 2   Montana
## 3 New Mexico
```

Rape Factor

Based on previous exercise I want to get **5 State**'s names with the lowest Rape ratio (Rape is provided by **USArrests** data set).

```
data("USArrests")

USArrests$State <- rownames(USArrests)
```

```

stanyWgWspPrzestepczpSciSeksualnejRoznaco <- USArrests %>%
  filter(State %in% stanyZPowierzchniaMiedzi50ka200k$State) %>% # filter the states based on previous
  arrange(Rape) # data set is Rape order, increasing

stanyWgWspPrzestepczpSciSeksualnejRoznaco %>%
  select(State) %>% # Selecting State variable
  head(5) # 5 states with the lowest Rape

```

```

##      State
## 1 North Dakota
## 2   Wisconsin
## 3       Iowa
## 4 South Dakota
## 5       Idaho

```

Working on **stanyZPowierzchniaMiedzi50ka200k** mutate new boolean variable with information if the State Income exceed 4600 or not.

```

stanyZPow50ka200k_filtr<- stanyZPowierzchniaMiedzi50ka200k %>%
  mutate(has_income = as.character(Income>4600)) # new variable with income grater than 4600 (boolean ty

```

Including Graphics

