DATA ANALYTICS AND MACHINE LEARNING WITH R

ASSIGNMENT 1

LUIS GUSTAVO NARDIN

INTERNET TECHNOLOGY

BRANDENBURG UNIVERSITY OF TECHNOLOGY

OVERVIEW

The goal of this assignment is to support the students to gain experience of using exploratory data analysis in a practical setting. The general objective is to summarize the dataset of The Fund for Peace about States fragility.

FUND FOR PEACE

*"The **Fund for Peace** (FFP) has been a world leader in developing practical tools and approaches for reducing conflict for over 60 years."*

"FFP is focused on understanding and addressing issues of violent conflict, state fragility, and security and human rights."

THE FRAGILE STATES INDEX

- based on a conflict assessment framework developed for assessing the vulnerability to states to collapse.
- FSI is a ranking of 178 countries across **12 indicators** of the risks and vulnerabilities faced by individual nations.
- The indicators provide a snapshot in time that can be measured against other snapshots in a time series to determine whether conditions are improving or worsening

ACTIVITIES

1. Combine the FSI datasets from 2006 to 2018 available at **http://fundforpeace.org/fsi/excel/** into a single output dataset
   - each column containing a header
   - fields separated with semi-colon
   - filter out all rows with NA values

ACTIVITIES

2. Generate a line plot of the Total value of the indicators for a sample of 10 countries over the years
3. Generate a histogram of an indicator of your preference for all countries showing with color the count of countries on each bin per year

ACTIVITIES

4. Generate a line plot of all of the indicators of a single country (NOTE: Use the function `melt` in package `reshape2` to re-shape the dataset for a suitable form, see next slide).

MELT

Suppose you have a dataset as follows and you need to make the make `c1` and `c2` factors to generate graphics

```
> data <- data.frame(country=c("Togo","Mauritania","Togo","Mauritania"),
                     year=c("2006","2006","2018","2018"),
                     c1=c(8.1,7.6),
                     c2=c(7.8,7.9))
> data
     country year  c1  c2
1       Togo 2006 8.1 7.8
2 Mauritania 2006 7.6 7.9
3       Togo 2018 8.1 7.8
4 Mauritania 2018 7.6 7.9
```

MELT

Use the `melt` function indicating the variables that are id and those that are measurements.

```
> melt(data, id.vars=c("country","year"), measure.vars=c("c1","c2"))
     country year variable value
1       Togo 2006       c1   8.1
2 Mauritania 2006       c1   7.6
3       Togo 2018       c1   8.1
4 Mauritania 2018       c1   7.6
5       Togo 2006       c2   7.8
6 Mauritania 2006       c2   7.9
7       Togo 2018       c2   7.8
8 Mauritania 2018       c2   7.9
```