

Exploratory Data Analysis and Visualization - MSc AIDA UoM

Professor George Evaggelidis

Academic Year: 2021-2022

You will use the `EconomistData` dataset. The data are from the Economist magazine and concern two indices:

- **HDI** = Human Development Index
- **CPI** = Corruption Perceptions Index

```
dat <- read_csv("~/EconomistData.csv") % if you are on igreed
dat <- read_csv("EconomistData.csv") % if you are running RStudio on your computer
glimpse(dat)
```

The HDI index takes values from 0 to 1, with 1 being the best, while the CPI index takes values from 0 to 10, with 10 being the best.

Acronym explanations:

- **MENA** = Middle East and North Africa
- **SSA** = Sub-Saharan Africa
- **East EU Cemt Asia** = Eastern Europe and Central Asian countries that have CEMT permits, where CEMT = Conférence Européenne des Ministres des Transports

Some Hints:

- Explore which Regions exist

```
unique(dat$Region)
```

- View the countries in the "EU W. Europe"

```
\texttt{dat \%>\% filter(Region=="EU W. Europe") \%>\% view() }
```

- Here's a comparison of the Regions with respect to the two indices

```
ggplot(dat2, aes(x = avgHDI, y = avgCPI, color=Region)) +
  geom_point()
```

- Achieve the same with dplyr functions

```
(dat3 <- dat \%>\%
  group_by(Region) \%>\% summarise(avgHDI = mean(HDI), avgCPI = mean(CPI)))

ggplot(dat3, aes(x = avgHDI, y = avgCPI, color=Region)) +
  geom_point()
```

To answer the following problems, consult the ggplot2 cheat sheet. There is no single answer as there are many ways to represent the requirements of each problem. Experiment with different plots and improvise. Submit this file with your answers to CoMPUs.

1. Provide a plot that describes the distribution of CPI.
2. Provide a plot that shows the number of countries per Region.
3. Provide a plot that shows the percentage of countries per Region.
4. Provide a plot that shows the average HDI per Region.
5. Provide a plot that shows the correlation of HDI and CPI for all countries.
6. Provide a set of plots (facet) that shows the correlation of HDI and CPI for all the countries in each Region.