Cross-correlation Analysis

```
library(tidyverse)
library(GGally)
```

Data preparation

Load and clean data.

```
sdg <- read.csv("2019_global_development_index.csv", stringsAsFactors = FALSE)

sdg <- as_tibble(sdg)

columns <- c("country", "id", "global_index_score", "global_index_score_1",
             "global_index_rank", "regional_score", "index_to_be_defined",
             "inverse_global_index_score", "sdr_mild_version", "sdr_medium_version",
             "sdr_heavy_version", "equality_population", "capacity",
             "historical_responsability", "present_responsability",
             "ponderated_index", "mcj")

colnames(sdg) <- columns

sdg %>% head()
```

```
## # A tibble: 6 x 17
##   country id    global_index_sc~ global_index_sc~ global_index_ra~
##   <chr>   <chr>            <dbl>            <dbl>            <int>
## 1 Centra~ CAF               39.1             39.1              162
## 2 Chad    TCD               42.8             42.8              161
## 3 Congo,~ COD               44.9             44.9              160
## 4 Nigeria NGA               46.4             46.4              159
## 5 Madaga~ MDG               46.7             46.7              158
## 6 Liberia LBR               48.2             48.2              157
## # ... with 12 more variables: regional_score <dbl>, index_to_be_defined <dbl>,
## #   inverse_global_index_score <dbl>, sdr_mild_version <dbl>,
## #   sdr_medium_version <dbl>, sdr_heavy_version <dbl>,
## #   equality_population <dbl>, capacity <dbl>, historical_responsability <dbl>,
## #   present_responsability <dbl>, ponderated_index <dbl>, mcj <dbl>
```

```
# Replace empty strings by NA
sdg[grepl("\"\"", sdg)] <- NA

# These are repeated columns
all(sdg$global_index_score == sdg$global_index_score_1)
```

```
## [1] TRUE
```

```
# Gather data for the correlation analysis
sdg_for_analysis <- sdg %>%
  select(-global_index_score, -global_index_score_1,
```

```
         -global_index_rank, -regional_score)

sdg_for_analysis_pivoted <- sdg_for_analysis %>%
  pivot_longer(-c("country", "id"), names_to = "indicator", values_to = "values")
```

Indicators for the analysis.

```
sdg_for_analysis_pivoted %>%
  group_by(indicator) %>%
  tally()
```

```
## # A tibble: 11 x 2
##    indicator                      n
##    <chr>                      <int>
##  1 capacity                     193
##  2 equality_population          193
##  3 historical_responsability    193
##  4 index_to_be_defined          193
##  5 inverse_global_index_score   193
##  6 mcj                          193
##  7 ponderated_index             193
##  8 present_responsability       193
##  9 sdr_heavy_version            193
## 10 sdr_medium_version           193
## 11 sdr_mild_version             193
```

Sumamry statistics

```
library(moments)

summary_1 <- sdg_for_analysis_pivoted %>%
  group_by(indicator) %>%
  summarise(
    number_of_observations = n(),
    missing_data = round(sum(is.na(values))/n(), 4)
  )

summary_2 <- sdg_for_analysis_pivoted %>%
  filter(!is.na(values)) %>%
  group_by(indicator) %>%
  summarise(
    mean = mean(values),
    skewness = skewness(values),
    kurtosis = kurtosis(values),
    minimum_value = min(values),
    maximum_value = max(values)
  )

summary_1 %>%
  left_join(summary_2)
```

```
## Joining, by = "indicator"
```

```
## # A tibble: 11 x 8
##    indicator number_of_obser~ missing_data   mean skewness kurtosis
```

```
##    <chr>                <int>        <dbl>  <dbl>    <dbl>    <dbl>
##  1 capacity               193       0.0415  0.541    2.56     10.8
##  2 equality~              193       0       0.518    8.34     76.7
##  3 historic~              193       0       0.518    1.80      6.24
##  4 index_to~              193       0      34.5      0.289     2.43
##  5 inverse_~              193       0       0.428    0.289     2.43
##  6 mcj                    193       0.917   3.16     2.20      6.20
##  7 ponderat~              193       0       0.522    7.03     63.9
##  8 present_~              193       0       0.518    2.08      8.08
##  9 sdr_heav~              193       0       0.518    9.05     96.7
## 10 sdr_medi~              193       0.145   0.606    7.74     72.2
## 11 sdr_mild~              193       0       0.518    8.51     82.6
## # ... with 2 more variables: minimum_value <dbl>, maximum_value <dbl>
```