

# R Skills Assessment Worksheet

Graduate Seminar - R Workshop  
Adee Weller & Eddy Yeung

## Objective

This worksheet is designed to test your ability to load, clean, analyze, and visualize real-world data in R. The tasks are divided into sections to assess different aspects of your proficiency. Please complete each section and submit your R script and outputs.

## Part 1: Data Loading & Exploration

### 1. Download the dataset

Use the *World Bank Governance Indicators (WGI)* dataset for the year 2020, focusing on governance scores by country. Download it from [here](#). Use the indicator “Rule of Law” as the focus for this exercise.

### 2. Load the dataset

Load the CSV file into R and assign the dataset to a variable called `governance_data`.

### 3. Explore the dataset

Use functions like `head()`, `summary()`, and `str()` to get a sense of the data structure.

- What is the range of values for the “Rule of Law” indicator?

### 4. Filter data

Filter out countries that do not have data for the “Rule of Law” indicator using `filter()` from `dplyr`.

- Report the number of countries left after filtering.

## Part 2: Data Cleaning

### 1. Handle missing values

Check for missing values in the dataset using `is.na()`. Remove rows with missing values from the dataset.

### 2. Rename columns

Rename the column containing country names to `Country` and the column with the “Rule of Law” score to `RuleOfLaw` using `rename()` from `dplyr`.

### 3. Create a new variable

Create a binary variable called `StrongRuleOfLaw`, which is 1 if the country’s Rule of Law score is above the median and 0 otherwise.

## Part 3: Descriptive Statistics & Visualization

### 1. Descriptive statistics

Calculate the mean, median, and standard deviation of the “Rule of Law” scores for the entire dataset and by region (if available).

- Provide a brief interpretation of these statistics. What do they tell us about global governance?
2. **Visualize the distribution**  
Create a histogram of the “Rule of Law” scores using `ggplot2`. Add appropriate labels and a title to the plot.
  3. **Boxplot by region**  
Create a boxplot of “Rule of Law” scores by region (if regional data is available) or by income group using `ggplot2`.

## Part 4: Regression Analysis

1. **Run a simple linear regression**  
Use `lm()` to regress “Rule of Law” (dependent variable) on a relevant independent variable such as “Control of Corruption” (another WGI indicator).
  - Report the coefficient, R-squared, and p-value. Interpret these results.
2. **Run a multiple regression**  
Add another variable, like GDP per capita (which you can find in another dataset), as an additional predictor and rerun the regression.
  - Report the coefficients, R-squared, and p-values. How do the results change?

## Part 5: Advanced Visualization

1. **Scatter plot with regression line**  
Create a scatter plot of “Rule of Law” vs. “Control of Corruption” and add a regression line using `ggplot2`.
2. **Facet grid by region**  
Create a facet grid that shows scatter plots of “Rule of Law” vs. “Control of Corruption” by region. Add regression lines to each plot.
  - Briefly explain any regional patterns that you observe.

## Submission Guidelines

Save your R script as `RSkills_Assessment.R`. Save your output visualizations and statistical summaries as a PDF report compiled using `LATEX`. Submit both the R script and the PDF report.

## Optional (for advanced students)

1. **Logistic regression**  
Use `glm()` to run a logistic regression predicting `StrongRuleOfLaw` (binary variable) based on “Control of Corruption” and “GDP per capita”. Report the coefficients, p-values, and model fit statistics (AIC).
2. **Clustering analysis**  
Use k-means clustering to group countries based on “Rule of Law” and “Control of Corruption” scores. How many clusters make sense based on the data? Visualize the clusters in a scatter plot.

## Feedback Questions

- How confident did you feel completing each section of the worksheet?
- Which sections were the most challenging for you?
- What specific topics would you like to cover in future sessions?