

# Homework 4

## Time Series Data Visualization

### Requirements

- How to read the requirements? **Carefully**
- File format: For this Homework, you are required to submit both R Markdown and PDF files with your answers and codes in it and .ipynb file for Python section. Make sure that script files work, there won't be any errors when it is run and represent the same information as PDF Under each question (not in comments) write the code along with your interpretations. **Be sure to put your name at the top of your assignment (in the YAML header in front of the author).**
- **No late homework will be accepted.**
- Submission: **You need to upload github links on Moodle as comments**
- Suggestion: Start with creating a framework for your homework, retype the tasks and only then start to solve the tasks.
- Rule of thumb: If the number of data points is greater than 50, **do not print the whole data.** Use subsets. Try to show all outputs (do not just store an object as a variable). Also, try to avoid using the same name for variables in the file.
- Cheating: The purpose of tasks is to check your knowledge (rather than the ability of thinking). Please, try to solve without googling every exercise. Try not to discuss with your classmates and work only on your file. **Any similarities, which can be considered as cheated, will not be graded.**
- Note: Keep the order of sub-tasks. Pay attention to titles, legends and axes of graphs.

Good luck!

The file bundesliga contain games played in german football league BUNDESLIGA, the variable descriptions:  
SEASON: from 1994-2019

LEAGUE: The name of the league: The first league of every country is included DATE: the date of the game HOMETEAM: The name of the home team AWAYTEAM: The name of the away team FTSC: Full Time score FTR: Full Time Result (H=Home Win, D=Draw, A=Away Win, “-”=unknown) FTHG: Full Time Home Team Goals FTAG: Full Time Away Team Goals FTTG: Full Time Total Goals

### Optional Objective for Homework

Your goal is to take the csv file, do data manipulation and transformations to get the data exactly as is bundesliga2.csv (Final table at the end of the season)

# Main Objective for Homework

You are tasked with producing high-level statistical visualizations of the Bundesliga and Bundesliga2 datasets using only core visualization libraries. Every insight must be backed by code, and every chart must stand on its own. This is not just about aesthetics — it's about data storytelling under constraints. Remember to use both datasets for your visualizations.

## Part 1: Trend Analysis (R or Python)

1. Analyse trend of goals per season. For example total goals per match, average goals per match. (Python or R)
2. Goal Distribution Per Season. Use appropriate type of graphs for goals per match, year-wise. Color-code by whether average is above or below 2.5 (over/under bet threshold).
3. Create line charts for each season. Visualize trend of goals for each team that played in that season. Highlight only Bayern Munchen with red color. Rest should be gray. Add appropriate title that will contain information about season and total scored goals. Add footnote mentioning total number of goals scored by Bayern Munchen for that season. Save all graphs in pdf.

## Part 2: Home Advantage Deconstructed (R or Python)

1. Create Heatmap of Home vs. Away Wins per Team per Season
2. Point Differential Density: Create visualizations that will show difference per team for home and away game wins.

## Part 3: (R or Python)

1. Team Trajectories and Volatility
  - Seasonal Position Trajectories
  - Line plots showing seasonal ranks for top 6 teams.
  - Annotate title-winning seasons.
2. Volatility Index
  - For each team, calculate standard deviation of final rank over all seasons.
  - Use a bar chart with conditional coloring (e.g., red = unstable, green = consistent).
  - Add text labels above each bar with exact values.

## Part 4: Rivalries & Big Match Patterns (R or Python)

1. Head-to-Head Matrix for Selected Rivalries
  - Select 5 key rivalries more info [click here](#) .
  - Create a facet grid of win/draw/loss bar charts per rivalry.
  - Annotate biggest win margins.
2. Upset Visualizer

- Define “upset” as a team >8 places below beating a top-5 team.
- Scatterplot of upsets: x-axis = rank difference, y-axis = goal difference.
- Encode team colors; highlight and label famous upsets

Note you can define top 5 by most points, most scored goals, less conceded goals.

## **Part 5: Overall performance (R and Python)**

- Define unique color for each team per season. For each season create horizontal bar plot using total number of points. Highlighting the winner with the unique color that you assigned to it. Save all graphs in pdf. (R)
- Redo the same task in python. But instead of total points use goal difference. Use same logic for colors as in first part. (Python)

## **Part 6. Monte Carlo simulation. (R or Python)**

Use Monte Carlo simulation to predict how many goals will Bayern Munchen score for next 10 seasons. Repeat the same for Bayer Leverkusen and Borussia Dortmund. Compare results using appropriate visualization technique.

For more about monte carlo simulation click [here](#)