*Please check the requirements for Innovationlab projects and fill out the project details below.*

# Checklist for project partner:

* Are the data are already collected and available?
* In case that data should be collected by the students (for example through web-scraping), are the data sources clearly defined? (e.g. “search some useful data on the internet” is not enough).
* Are the students allowed to work with the data on their machines?
* Are all the data tables and columns documented?
* Is the output of a successful project clearly defined?
* Is the project doable for a group of 4-5 students within 3 - 4 months? (we will help you define a reasonable scope)
* The results will be presented to the other students and the supervisors. The project partner allows this without requiring the other students to sign NDAs.
* The project partner allows an abstract of the project to be published on the InnovationLab website. The name of project partner can be removed if requested.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Election analyses - Creation of backend functionality for automated data gathering and handling

# Project Background

**KOALA database setup**

In our KOALA (Koalitions-Analyse) project the Statistical Consulting Unit (StaBLab) at the Department of Statistics, LMU, uses published election polls to estimate probabilities for specific events (if the election was held today), e.g. “Will two parties obtain enough votes to form a governing coalition (joint majority)?”. Such event probabilities are better suited to communicate election polls and their underlying uncertainty to the public than only the observed voter shares. Results are communicated to the public on koala.stat.uni-muenchen.de and on Twitter (@KOALA\_LMU).

What we are still missing is a database to save observed opinion polls, the estimated probabilities and intermediate results, including a public API, usable for media access.

The primary aim of this project is to set up a database (based on a consistent data structure) and to build a public API. The database should be accompanied by an R package with functions to retrieve, edit and save new data in the database, with complete access for admins and read-only access for media partners.

**Automated election data gathering**

Another recurring issue is the availability, but lack of accessibility of election result in the immediate post-election period. This lack of accessibility hinders post-election analyses. The StaBLab together with LMU political science cooperated in a post-election estimation of voter transition in 2013 and 2018, also strongly relying on the availability of such information.

In this part of the project an automated information gathering algorithm should be elaborated. The aim of this algorithm is to gather the published election results at local level websites (e.g. of towns or villages) and combine them to one useable dataset. Different ways to present the data should be handled by the algorithm in a flexible manner, also to ensure future adaptations of the algorithm to other contexts. For this project itself the focus is set on election data published on the websites from Bavarian towns and villages.

# Milestones

1. Set up a modern database with a public API
2. Build an R package for database access
3. Discuss the applicability of one general data gathering algorithm
4. Implement said algorithm

# Data

As we work with R all data are available as R data.frames or lists. Data size is easily manageable, <100MB.