Exercise A

Load the `Boston` dataset from the `MASS` package. Think of a simple (linear) model aiming at predicting property prices involving 4–5 covariates (or interactions between some of them).

- Create a function that takes your dependent variable and the covariates as inputs, and return a list with:
  - OLS point estimates for the intercept, slope parameters, and the error variance.
  - Suitable test statistics with corresponding p-values for the relevant coefficients.
  - Intervals of the coefficients for a confidence level of 95%.

Exercise B

Come up with some network of interest, vaguely related to some real-world example (describe very briefly), with *at least* six agents and ten edges between them.

1. Draw a graph of the network; create the adjacency matrix in `R`.
2. Who are the most and least central agents in the network? Name, explain, and try to quantify different notions of centrality.
   - How would centralities change if you considered a row-normalized network instead?
   - How would the network change if you removed or added a specific agent?
3. Simulate some agent characteristic based on a standard Normal distribution; use this characteristic to simulate responses following a linear-in-means model with your network.
   - Repeat this a couple of times, and compare estimates of a standard linear model ($y_i = x_\beta + \varepsilon_i$) with the true values that you used to simulate the data.

Exercise C

Download a suitable shapefile for NUTS2 regions (from here or using `R` directly) and some dataset of interest at the same level of aggregation (e.g. from here). Install and load the `sf` and `ggplot2` packages together with their dependencies.

- Read in the shapefile, and find out what projection and CRS the file uses. Map the data to use another projection and/or CRS of your choosing.
- Merge the shapefile and the chosen dataset. Create **two meaningful** visualizations of the chosen dataset using different scales (e.g. continuous versus discrete scaling of the data).

- Briefly explain what two conceptually different ways there are to store visualizations, and name two formats each. Which ones are more appropriate for which types of visualizations and why? Use the one you think is more appropriate to store your visualizations.

Exercise D

Install and load the `tmap` and `spDataLarge` packages (available from GitHub). Load and review the `pol_pres15` dataset on the Polish Presidential election in 2015 (see `?pol_pres15`). Create **three different, insightful visualizations** of the underlying data.

- One visualization should compare the support for Komorowski and Duda.
- One visualization should investigate possible issues with postal voting envelopes.
- At least one visualization should use the `tmap` package. You can use the following example (`tm_facets` allows plotting multiple variables) as a point of departure:

```
tm_shape(pol_pres15) + tm_facets(free.scales = FALSE) +
    tm_borders(lwd = 0.5, alpha = 0.4) + tm_fill("types")
```