

Assignment 1 - Group 8

Elia Di Gregorio and Robert Auerbach

2024-03-28

Quick note

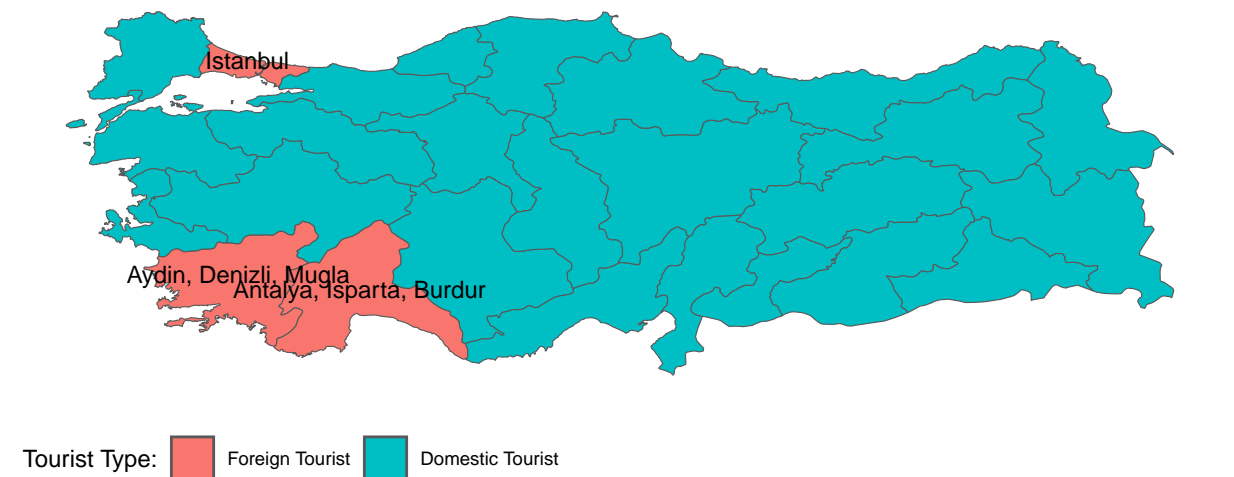
Super cool trick for the R_Markdown is to have saved the workspace of your basic R_file somewhere in the project and then load it directly. In this way you have all the necessary objects and plots from that basic R_file, within your R_Markdown_file, too. There is no need to repeat and rerun the code to plot something. You can just “call” it :) Of course for those exercises where the focus was developing a code/function it then make sense to include the whole code.

```
source("../code/__packages.R")  
load("../WS_CD.RData")
```

```
TR_plot_2
```

Distribution of Tourist Overnight–Stay by Origin

Türkiye – NUTS2 Level



Source: Eurostat

Exercise C In this exercise, we will work with spatial projections of Türkiye’s NUTS-2 regions. Despite not being a member of the European Union, it has adopted the Nomenclature of Territorial Units for Statistics (NUTS) since 2002.

Different projections

Türkiye counts 26 subregions, and we accessed its spatial data directly from the GISCO source via `get_eurostat_spatial()` function. By default this function loads the EPSG-4326 projection of the map, corresponding to the World Geodetic System 1984 ensemble (WGS84). It is based on a geocentric datum, meaning it defines the Earth's shape as an ellipsoid (a flattened sphere) rather than a perfect sphere.

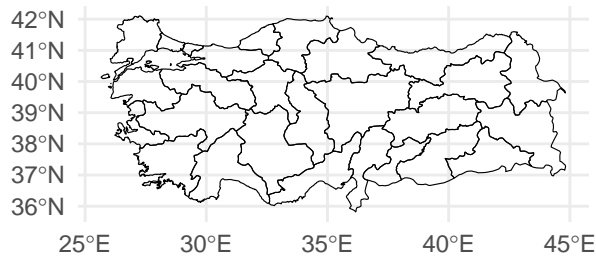
```
st_crs(Türkiye)
```

```
## Coordinate Reference System:
##   User input: EPSG:4326
##   wkt:
##   GEOGCRS["WGS 84",
##     ENSEMBLE["World Geodetic System 1984 ensemble",
##       MEMBER["World Geodetic System 1984 (Transit)"],
##       MEMBER["World Geodetic System 1984 (G730)"],
##       MEMBER["World Geodetic System 1984 (G873)"],
##       MEMBER["World Geodetic System 1984 (G1150)"],
##       MEMBER["World Geodetic System 1984 (G1674)"],
##       MEMBER["World Geodetic System 1984 (G1762)"],
##       MEMBER["World Geodetic System 1984 (G2139)"],
##       ELLIPSOID["WGS 84",6378137,298.257223563,
##         LENGTHUNIT["metre",1]],
##       ENSEMBLEACCURACY[2.0]],
##     PRIMEM["Greenwich",0,
##       ANGLEUNIT["degree",0.0174532925199433]],
##     CS[ellipsoidal,2],
##       AXIS["geodetic latitude (Lat)",north,
##         ORDER[1],
##         ANGLEUNIT["degree",0.0174532925199433]],
##       AXIS["geodetic longitude (Lon)",east,
##         ORDER[2],
##         ANGLEUNIT["degree",0.0174532925199433]],
##     USAGE[
##       SCOPE["Horizontal component of 3D system."],
##       AREA["World."],
##       BBOX[-90,-180,90,180]],
##     ID["EPSG",4326]]
```

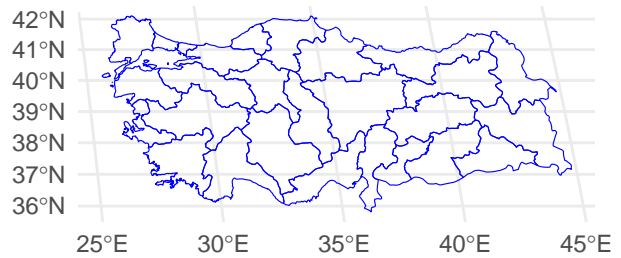
To use another projection and/or CRS we employed the `st_transform()` function and recurred to the Equal Earth Projection and to the Lambert Azimuthal Equal Area Projection. Both preserve the relative sizes of areas on the Earth's surface. This means that areas on the map are represented accurately in relation to each other in terms of size, making it suitable for thematic mapping and spatial analysis. However the former is classified as a pseudo-cylindrical projection because it projects the Earth's surface onto a cylinder and then unwraps the cylinder to form a rectangular map, whereas the latter projects the Earth's surface onto a plane tangent to a specific point (the center of the projection). The difference in projections is highlighted by the following map:

```
plot(TR_plot_1)
```

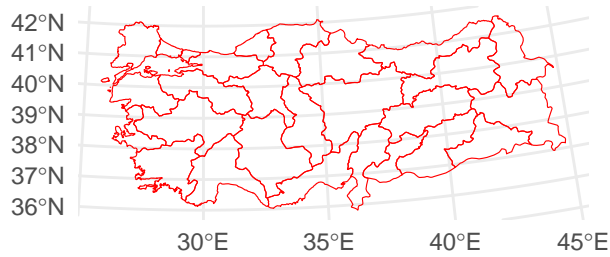
WGS 84 Projection



Equal Earth Projection



Lambert Azimuthal Projection



Source: Eurostat

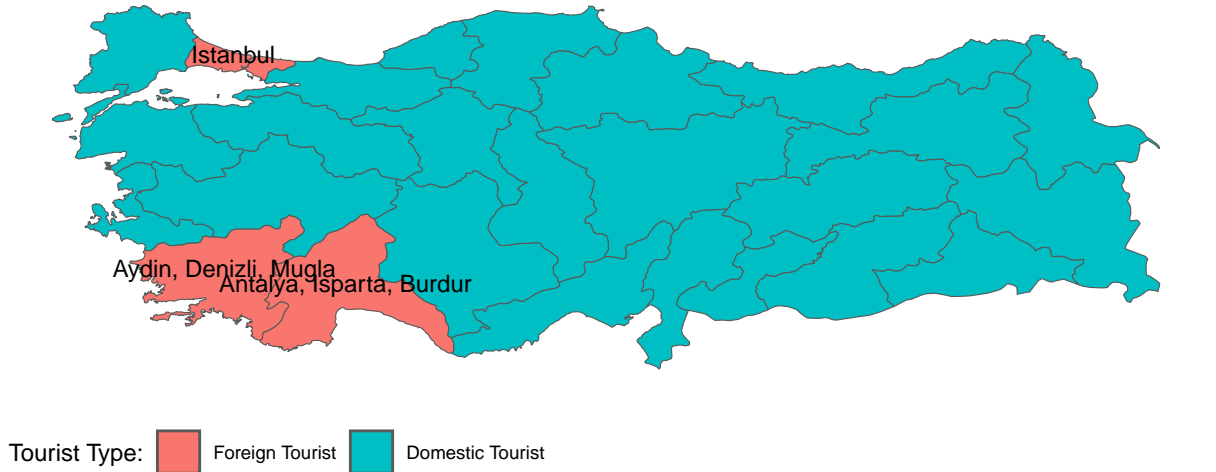
Map and Dataset The dataset used was `tgs00111`: Nights spent at tourist accommodation establishments by NUTS 2 regions (here the Metadata for consultation). The dataset covers internal tourism, in other words tourism flows within the country (domestic tourism) or from abroad to destinations in the country (inbound tourism) for the year 2022. For the sake of the exercise we further enhanced the dataset with two additional columns with the share of domestic and foreign tourist overnights (continuous scale), and one column for a factor variable based on the condition that if the domestic tourism share of overnight stay is greater than 50%, the tourist is labeled as “Domestic Tourist”; otherwise, they are labeled as “Foreign Tourist”.

By plotting the latter variable, we were able to group together regions based on whether the majority of overnight stay in hotels were by domestic tourists or foreign tourists.

```
plot(TR_plot_2)
```

Distribution of Tourist Overnight–Stay by Origin

Türkiye – NUTS2 Level



Source: Eurostat

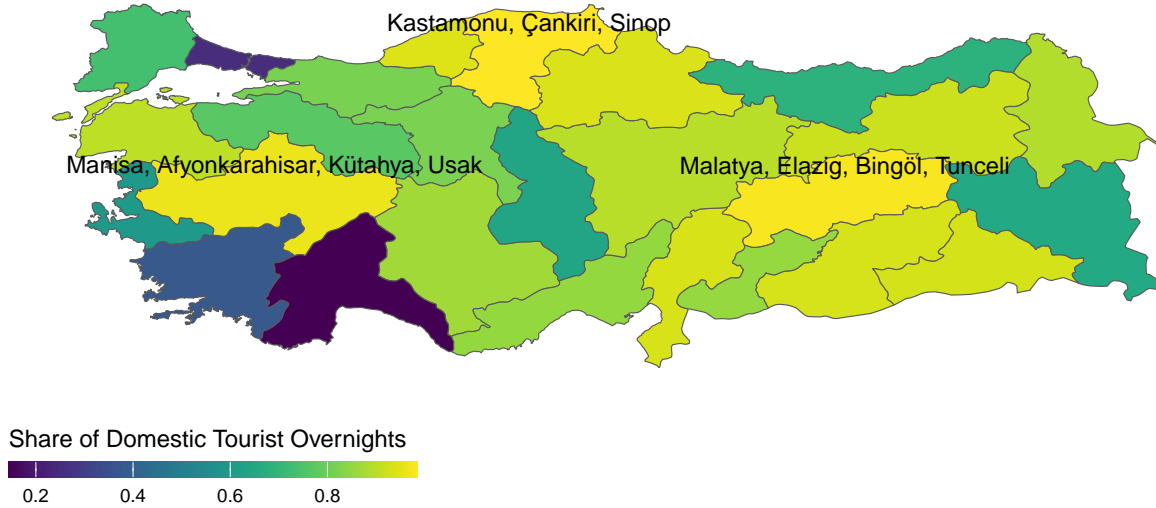
Not surprisingly, regions with famous “instagrammable” destinations like Istanbul or Antalya and Burdur had a majority of foreign tourists staying overnight in hotels, while the rest of the country registered a higher domestic tourists inflow, suggesting a limited international touristic network in Türkiye.

The plot for the continuous scale variable focused more on the domestic dimension of tourism and represented the share of domestic tourist overnight stays in different Turkish regions.

```
plot(TR_plot_3)
```

Distribution of Domestic Tourist Overnight-stay

Türkiye – NUTS2 Level



Source: Eurostat

Besides stressing the point made by the previous map, this time we get more insights on the distribution of Domestic tourists in Türkiye. To further enrich the plot, we also decided to highlight the names of those 3 regions with beyond 95% of domestic tourists overnight stays.

Storing visualizations

There are two conceptually different ways to store visualizations: raster-based and vector-based formats.

Raster-based formats like PNG and JPEG store images as grids of pixels, making them suitable for complex color gradients and detailed images. However, they are resolution-dependent, which means they can lose quality when scaled up.

On the other hand, vector-based formats such as SVG and PDF store image data using mathematical formulas to define shapes, lines, and colors. This makes them ideal for visualizations with geometric shapes, charts, maps, and illustrations where scalability and high-quality printing are crucial. Unlike raster formats, vector graphics are resolution-independent and can be scaled without loss of quality.

For visualizations created using R and ggplot2, which inherently produce vector graphics, it is recommended to save them in vector-based formats like SVG or PDF. These formats maintain sharpness and clarity when scaled to any size, making them suitable for presentations, printing, and high-resolution displays. SVG is particularly useful for web-based graphics and interactive visualizations, while PDF is excellent for high-quality printing and cross-platform compatibility.

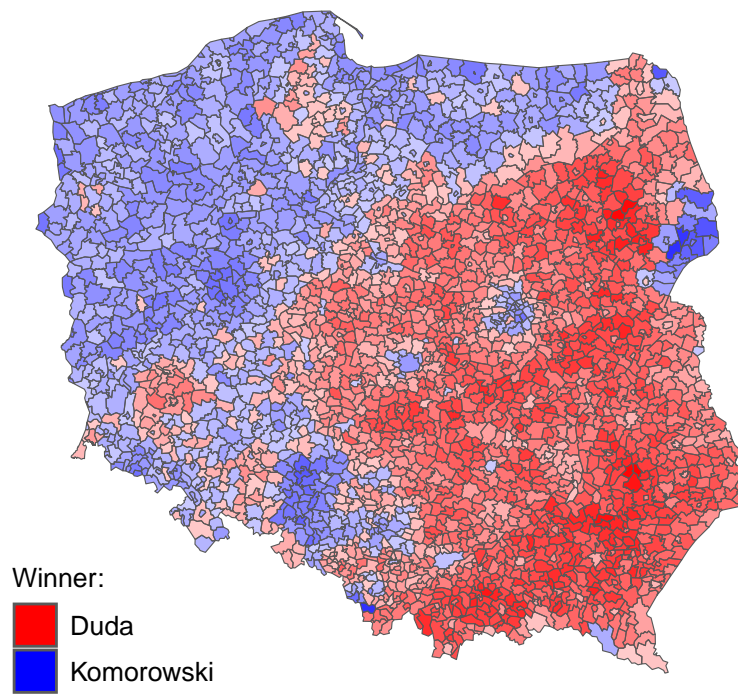
```
# Save the plot as SVG
#ggsave("TR_plot_2.svg", plot = TR_plot_2, device = "svg")
```

Exercise D

Comparison of support for Komorowski and Duda

```
plot(PL_plot_1)
```

2015 Polish Presidential Election: Duda vs. Komorowski
Poland – Municipality Level



Source: PKW

The provided R code generates a thematic map representing the outcomes of the second round of the 2015 Polish presidential elections at the municipality level. The data is grouped based on whether the winning candidate in each municipality was Duda or Komorowski. The plot uses different shades of red or blue (representing Duda or Komorowski, respectively) to indicate varying levels of majority support for the respective candidate.

The transparency of the colors also varies to reflect the degree of majority/support for the winning candidate in each municipality. Darker shades indicate a higher level of majority/support, while lighter shades represent a lower level of majority/support. This additional dimension helps visually emphasize areas where the winning candidate had a more significant lead in terms of voter share.

Postal voting envelopes anomalies

Three types of anomalies are identified:

- **anomaly_invalid**: Checks if there are more invalid voting papers than postal voting envelopes received in either round of voting.

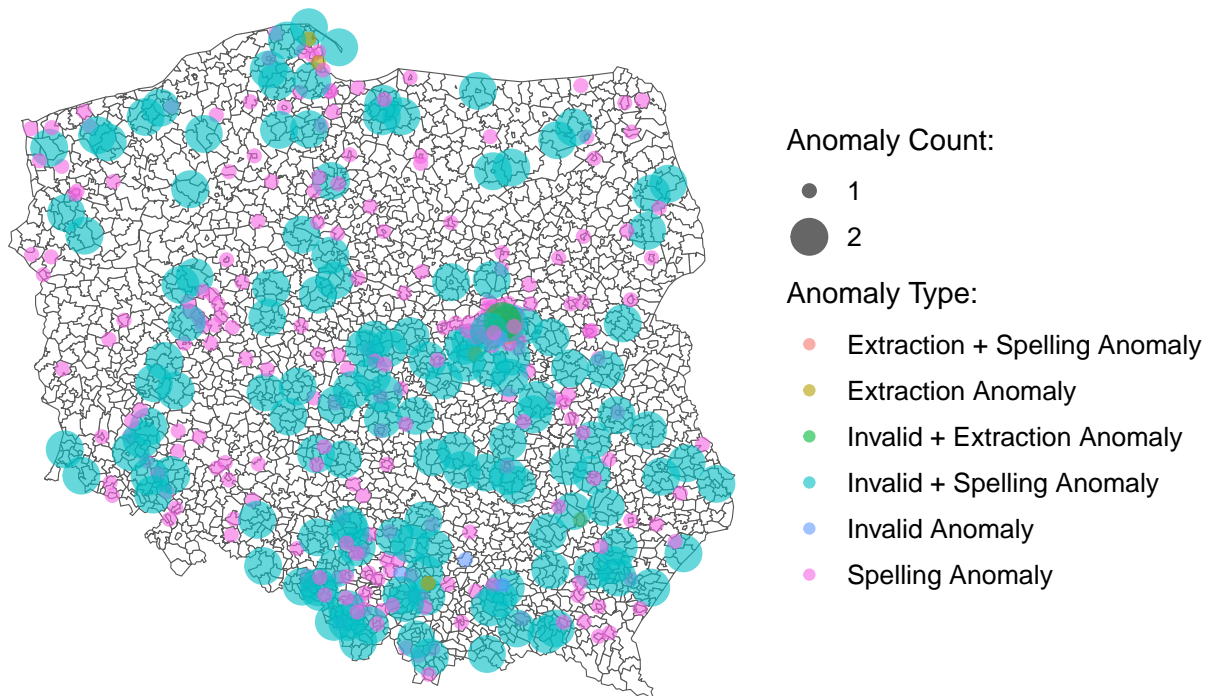
- **anomaly_spelling:** Identifies anomalies related to discrepancies in the number of invalid voting papers and specific errors on postal voting envelopes, such as missing declarations, signatures, voting envelopes, or signs of envelope opening.
- **anomaly_count:** Detects anomalies in the count of voting envelopes placed in the ballot box versus the number of voting papers taken from envelopes in either round of voting.

PL_plot_2

Warning: Using size for a discrete variable is not advised.

2015 Polish Presidential Election: anomalies in PVE

Poland – Municipality Level



Data source: PKW

Anomalies are represented as points (`geom_point()`) on the map, with each point's color indicating the type of anomaly and size representing the how many anomalies types were overall detected. In this way, one can highlight areas where specific types of anomalies or their combinations were observed more frequently.

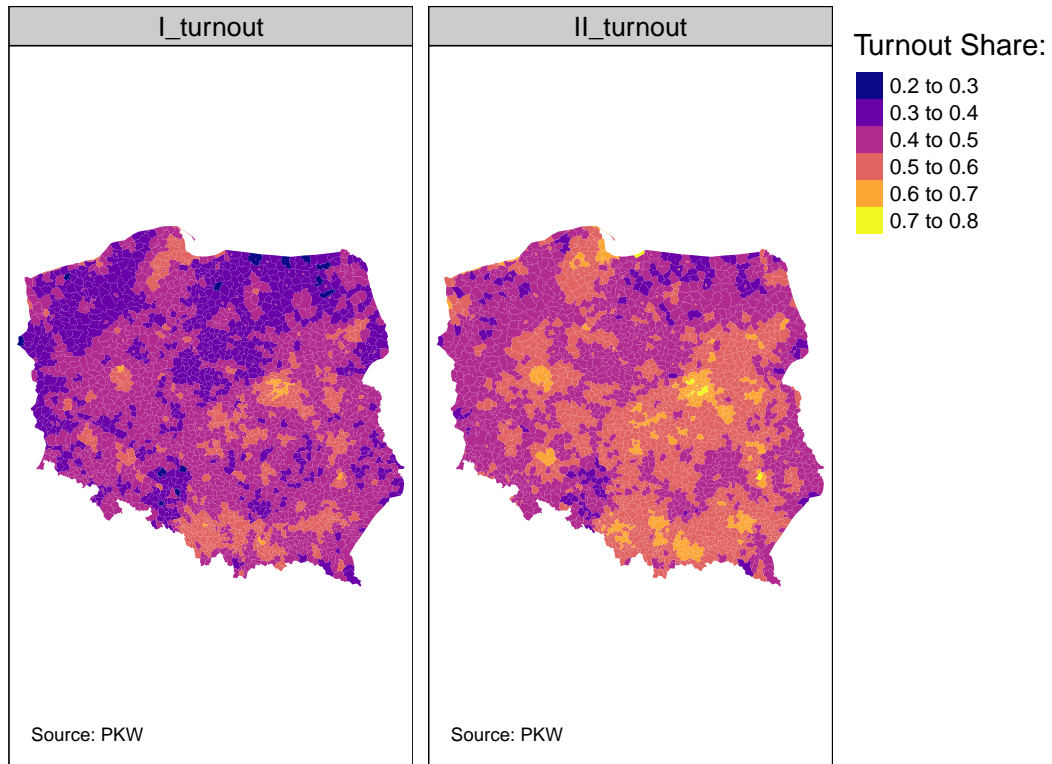
By visually exploring anomalies in the handling of voting materials, it appears that this was a widespread problem, involving several municipalities, although no clear pattern can be detected.

Notice that out of the three anomalies, the one about spelling is the “weak” one. Despite postal voting envelopes were return with a declaration, and a sealed voting envelope, the count of invalid PVE is still higher. It is not clear though, whether this is due to some further unrecorded issues (e.g. the misspelling of the candidate's name, thus the label for this variable). Further material, information or context should be validate the theory.

Turnout for each election round

For the final visualization, employing `tm_shape()` function, Turnout share in the first and second round of the presidential election was plotted. We made use of the `tm_facets` specification to combine in a single plot both heatmaps.

Turnout Comparison between I and II round



Overall, it appears that in the second round there was a higher public involvement. In rural areas this was the less stronger than in the large urban areas, where participation peaked as far as reaching almost 80%.