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 :)

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

```r
TR_plot_2
```


Distribution of Domestic Tourist Overnight–stay
Turkey – NUTS2 Level

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, Türkiye has adopted the Nomenclature of Territorial Units for Statistics (NUTS) since 2002. The country 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:
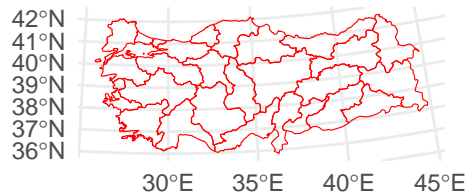
```
plot(combined_plot)
```

WGS 84 Projection
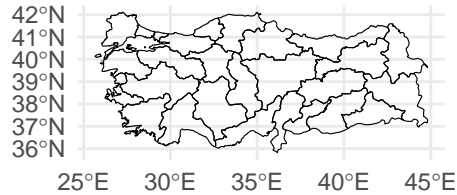
Equal Earth Projection
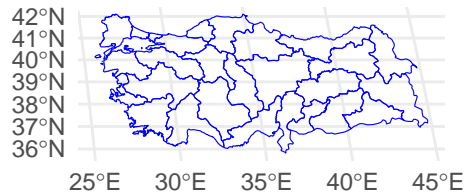


Lambert Azimuthal Equal Area Projection

Source: Eurostat

For the second part of the Exercise we decided to use the dataset `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_1)
```

Distribution of Tourist Overnight–Stay by Origin
Turkey – NUTS2 Level

Istanbul

Aydin, Denizli, Muğla
Antalya, Isparta, Burdur
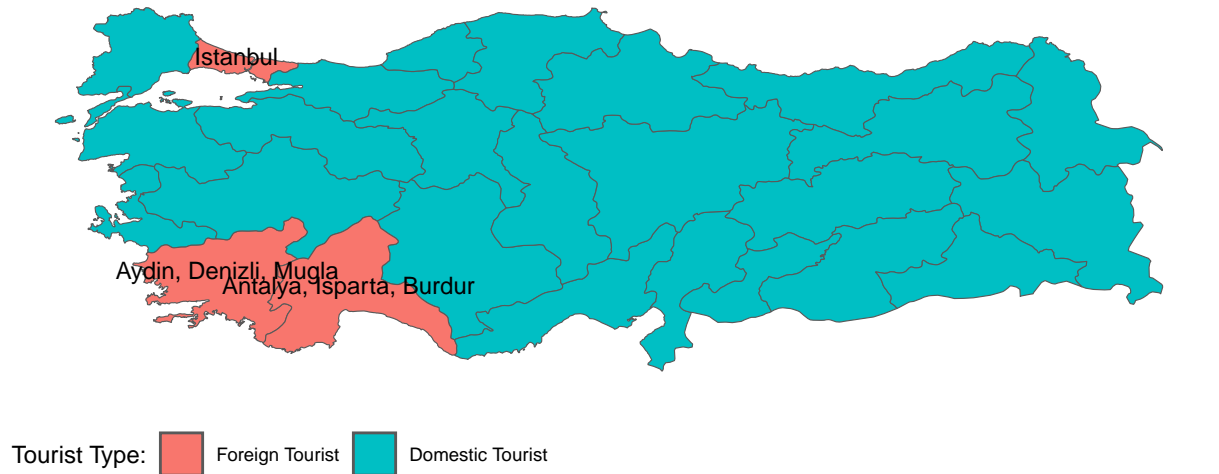
Tourist Type: Foreign Tourist   Domestic Tourist

Not surprisingly, regions with traditional "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 very 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_2)
```

Distribution of Domestic Tourist Overnight–stay
Turkey – NUTS2 Level

Share of Domestic Tourist Overnights

0.2 0.4 0.6 0.8

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 10 regions with beyond 90% of domestic tourists overnight stays.

Concerning the last question of the exercise, 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