

CENSUS Graphs

Import libraries

```
library(dplyr)
library(geojsonio)
library(broom)
library(ggplot2)
library(cowplot)
library(hrbrthemes)
library(treemapify)
```

Import Qualità della vità 2020 dataset

```
df = read.csv("https://raw.githubusercontent.com/IlSole24ORE/QDV/main/20201214_QDV2020_001.csv")
```

Internet >= 100 Mbit/s - abbonamenti

Import subset

```
df_inter = df[df$INDICATORE == "Internet â‰¥ 100 Mbit/s - abbonamenti",]
```

Retrieve NUTS2 code from NUTS3 (deleting the last character)

```
df_inter$REGION = gsub('.{1}$', '', df_inter$CODICE.NUTS.3.2021)
```

Substitute NUTS2 cose with region name

```
df_inter[df_inter$REGION == "ITC1",]$REGION <- "Piemonte"
df_inter[df_inter$REGION == "ITC2",]$REGION <- "Valle d'Aosta/Vallée d'Aoste"
df_inter[df_inter$REGION == "ITC3",]$REGION <- "Liguria"
df_inter[df_inter$REGION == "ITC4",]$REGION <- "Lombardia"
df_inter[df_inter$REGION == "ITH1",]$REGION <- "Trentino-Alto Adige/Südtirol"
df_inter[df_inter$REGION == "ITH2",]$REGION <- "Trentino-Alto Adige/Südtirol"
df_inter[df_inter$REGION == "ITH3",]$REGION <- "Veneto"
df_inter[df_inter$REGION == "ITH4",]$REGION <- "Friuli-Venezia Giulia"
df_inter[df_inter$REGION == "ITH5",]$REGION <- "Emilia-Romagna"
df_inter[df_inter$REGION == "ITI1",]$REGION <- "Toscana"
df_inter[df_inter$REGION == "ITI2",]$REGION <- "Umbria"
df_inter[df_inter$REGION == "ITI3",]$REGION <- "Marche"
df_inter[df_inter$REGION == "ITI4",]$REGION <- "Lazio"
df_inter[df_inter$REGION == "ITF1",]$REGION <- "Abruzzo"
```

```
df_inter[df_inter$REGION == "ITF2",]$REGION <- "Molise"
df_inter[df_inter$REGION == "ITF3",]$REGION <- "Campania"
df_inter[df_inter$REGION == "ITF4",]$REGION <- "Puglia"
df_inter[df_inter$REGION == "ITF5",]$REGION <- "Basilicata"
df_inter[df_inter$REGION == "ITF6",]$REGION <- "Calabria"
df_inter[df_inter$REGION == "ITG1",]$REGION <- "Sicilia"
df_inter[df_inter$REGION == "ITG2",]$REGION <- "Sardegna"
```

Calculate the mean of value for each region

```
means <- aggregate(VALORE ~ REGION, data = df_inter, FUN = mean)
```

Add column with the means to the dataset

```
df_inter = df_inter %>%
  right_join(. , means, by=c("REGION"="REGION"))
```

Import geojson for choropleth map graph

```
spdf = geojson_read("/Users/elypa/limits_IT_regions.geojson", what = "sp")
spdf_fortified = tidy(spdf, region = "reg_name")
```

Join the geojson with the dataset

```
spdf_fortified = spdf_fortified %>%
  left_join(. , df_inter, by=c("id"="REGION"))
```

Plot the choropleth map graph

```
# Plot the result
choropleth_ita_tot_GDP = ggplot() +

# pass info on the map
  geom_polygon(data = spdf_fortified, aes(fill = `VALORE.y`, x = long, y = lat, group = group), color="black")

# customize theme and title
  labs(title="Internet >= 100 Mbit/s - abbonamenti",
        subtitle="Accessi broadband - In % sulla popolazione residente") +
  theme_void() +
  theme(
    plot.title = element_text(size= 30, hjust=0.5, color = "black", margin = margin(b = -0.1, t = 0.4, l = 0.5, r = 0.5)),
    plot.subtitle = element_text(size= 17, hjust=0.5, color = "black", margin = margin(b = -0.1, t = 0.4, l = 0.5, r = 0.5)),
    plot.background = element_rect(fill = "#fafafa", color = NA),
```

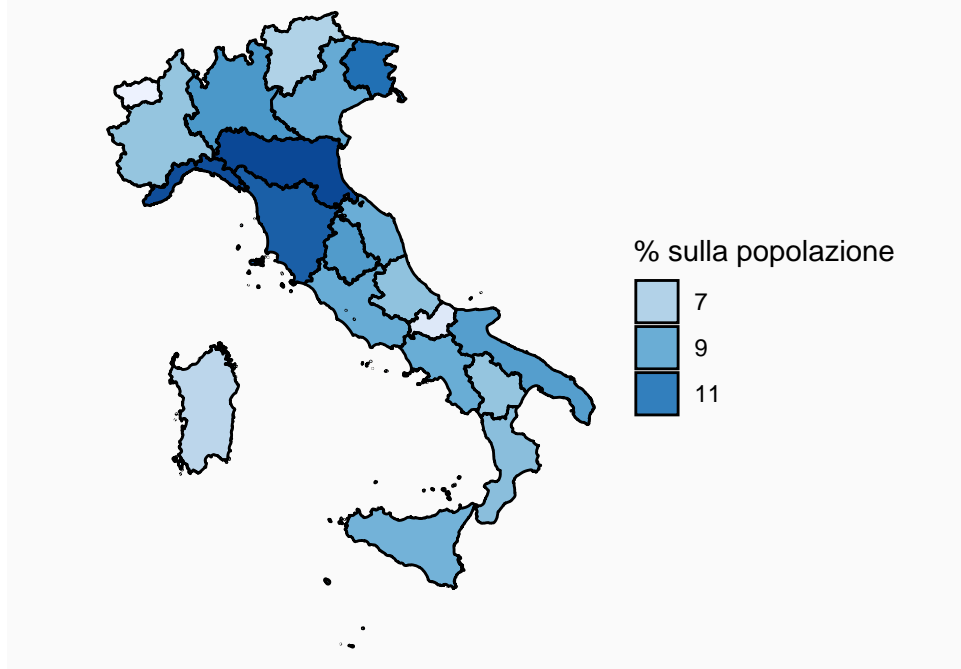
```

plot.margin = unit(c(0, 1, 0, 1), "cm")
) +
guides(fill=guide_legend("% sulla popolazione")) +
coord_map() +
scale_fill_distiller(guide = FALSE, direction = 1, limits=c(5.002,12.900))

choropleth_ita_tot_GDP

```

net >= 100 Mbit/s – abbonamenti cessi broadband – In % sulla popolazione residente



Prezzo medio di vendita delle case

Import subset

```
df_case = df[df$INDICATORE == "Prezzo medio di vendita delle case",]
```

Retrieve NUTS2 code from NUTS3 (deleting the last character)

```
df_case$REGION = gsub('.{1}$', '', df_case$CODICE.NUTS.3.2021)
```

Substitute NUTS2 cose with region name

```
df_case[df_case$REGION == "ITC1",]$REGION <- "Piemonte"
df_case[df_case$REGION == "ITC2",]$REGION <- "Valle d'Aosta/Vallée d'Aoste"
df_case[df_case$REGION == "ITC3",]$REGION <- "Liguria"
df_case[df_case$REGION == "ITC4",]$REGION <- "Lombardia"
df_case[df_case$REGION == "ITH1",]$REGION <- "Trentino-Alto Adige/Südtirol"
df_case[df_case$REGION == "ITH2",]$REGION <- "Trentino-Alto Adige/Südtirol"
df_case[df_case$REGION == "ITH3",]$REGION <- "Veneto"
df_case[df_case$REGION == "ITH4",]$REGION <- "Friuli-Venezia Giulia"
df_case[df_case$REGION == "ITH5",]$REGION <- "Emilia-Romagna"
df_case[df_case$REGION == "ITI1",]$REGION <- "Toscana"
df_case[df_case$REGION == "ITI2",]$REGION <- "Umbria"
df_case[df_case$REGION == "ITI3",]$REGION <- "Marche"
df_case[df_case$REGION == "ITI4",]$REGION <- "Lazio"
df_case[df_case$REGION == "ITF1",]$REGION <- "Abruzzo"
df_case[df_case$REGION == "ITF2",]$REGION <- "Molise"
df_case[df_case$REGION == "ITF3",]$REGION <- "Campania"
df_case[df_case$REGION == "ITF4",]$REGION <- "Puglia"
df_case[df_case$REGION == "ITF5",]$REGION <- "Basilicata"
df_case[df_case$REGION == "ITF6",]$REGION <- "Calabria"
df_case[df_case$REGION == "ITG1",]$REGION <- "Sicilia"
df_case[df_case$REGION == "ITG2",]$REGION <- "Sardegna"
```

Calculate the mean of value for each region

```
means <- aggregate(VALORE ~ REGION, data = df_case, FUN = mean)
```

Add column with the means to the dataset

```
df_case = df_case %>%
  right_join(. , means, by=c("REGION"="REGION"))
```

Import geojson for choropleth map graph

```
spdf = geojson_read("/Users/elypa/limits_IT_regions.geojson", what = "sp")
spdf_fortified = tidy(spdf, region = "reg_name")
```

Join the geojson with the dataset

```
spdf_fortified = spdf_fortified %>%
  left_join(. , df_case, by=c("id"="REGION"))
```

Plot the choropleth map graph

```
# Plot the result
choropleth_ita_tot_GDP = ggplot() +

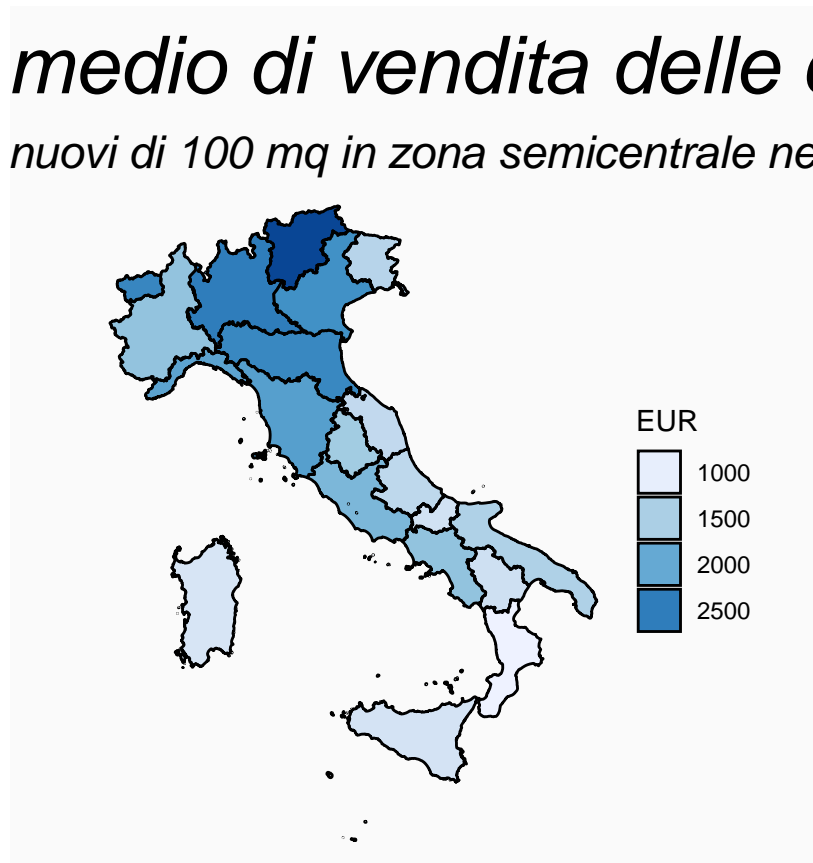
# pass info on the map
geom_polygon(data = spdf_fortified, aes(fill = `VALORE.y`, x = long, y = lat, group = group), color="black")

# customize theme and title
labs(title="Prezzo medio di vendita delle case",
      subtitle="Per appartamenti nuovi di 100 mq in zona semicentrale nei capoluoghi") +
theme_void() +
theme(
  plot.title = element_text(size= 30, hjust=0.5, color = "black", margin = margin(b = -0.1, t = 0.4, l = 0, r = 0)),
  plot.subtitle = element_text(size= 17, hjust=0.5, color = "black", margin = margin(b = -0.1, t = 0.4, l = 0, r = 0)),
  plot.background = element_rect(fill = "#fafafa", color = NA),
  plot.margin = unit(c(0, 1, 0, 1), "cm")
) +
guides(fill=guide_legend("EUR")) +
coord_map() +
scale_fill_distiller(guide = FALSE, direction = 1, limits=c(930,2960))

choropleth_ita_tot_GDP
```

Prezzo medio di vendita delle case

Per appartamenti nuovi di 100 mq in zona semicentrale nei capoluoghi



Spazio abitativo medio

Import subset

```
df_spazio = df[df$INDICATORE == "Spazio abitativo medio",]
```

Retrieve NUTS2 code from NUTS3 (deleting the last character)

```
df_spazio$REGION = gsub('.{1}$', '', df_spazio$CODICE.NUTS.3.2021)
```

Substitute NUTS2 cose with region name

```
df_spazio[df_spazio$REGION == "ITC1",]$REGION <- "Piemonte"
df_spazio[df_spazio$REGION == "ITC2",]$REGION <- "Valle d'Aosta/Vallée d'Aoste"
df_spazio[df_spazio$REGION == "ITC3",]$REGION <- "Liguria"
df_spazio[df_spazio$REGION == "ITC4",]$REGION <- "Lombardia"
df_spazio[df_spazio$REGION == "ITH1",]$REGION <- "Trentino-Alto Adige/Südtirol"
df_spazio[df_spazio$REGION == "ITH2",]$REGION <- "Trentino-Alto Adige/Südtirol"
df_spazio[df_spazio$REGION == "ITH3",]$REGION <- "Veneto"
df_spazio[df_spazio$REGION == "ITH4",]$REGION <- "Friuli-Venezia Giulia"
df_spazio[df_spazio$REGION == "ITH5",]$REGION <- "Emilia-Romagna"
df_spazio[df_spazio$REGION == "ITI1",]$REGION <- "Toscana"
df_spazio[df_spazio$REGION == "ITI2",]$REGION <- "Umbria"
df_spazio[df_spazio$REGION == "ITI3",]$REGION <- "Marche"
df_spazio[df_spazio$REGION == "ITI4",]$REGION <- "Lazio"
df_spazio[df_spazio$REGION == "ITF1",]$REGION <- "Abruzzo"
df_spazio[df_spazio$REGION == "ITF2",]$REGION <- "Molise"
df_spazio[df_spazio$REGION == "ITF3",]$REGION <- "Campania"
df_spazio[df_spazio$REGION == "ITF4",]$REGION <- "Puglia"
df_spazio[df_spazio$REGION == "ITF5",]$REGION <- "Basilicata"
df_spazio[df_spazio$REGION == "ITF6",]$REGION <- "Calabria"
df_spazio[df_spazio$REGION == "ITG1",]$REGION <- "Sicilia"
df_spazio[df_spazio$REGION == "ITG2",]$REGION <- "Sardegna"
```

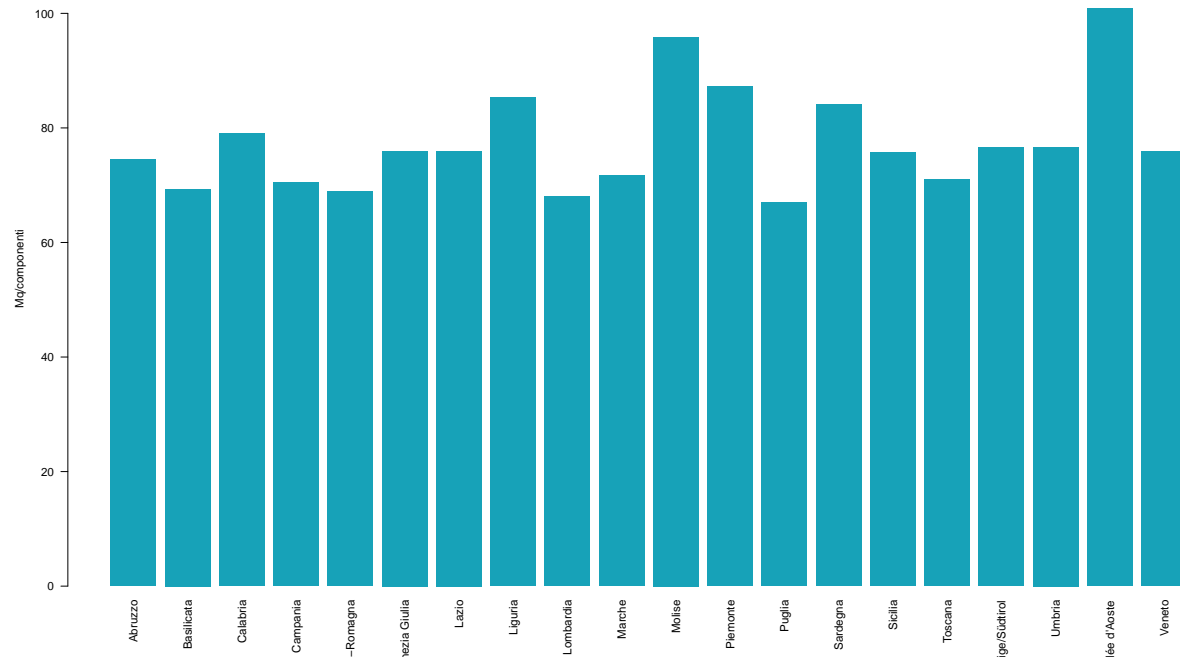
Calculate the mean of value for each region and store in a new datasets

```
means <- aggregate(VALORE ~ REGION, data = df_spazio, FUN = mean)
```

Barplot

```
# par(bg = '#fafafa-')
# par(mar = c(12, 5, 5, 5))
my_bar <- barplot(means$VALORE, border=F, names.arg=means$REGION,
                  las=2,
```

```
col= "#17a2b8",
ylim=c(0,110.88) ,
ylab="Mq/componenti",
main="" )
```



Tree map

```
means$mq <- means$VALORE

ggplot(means, aes(area = mq, fill = mq, label = REGION)) +
  geom_treemap() +
  geom_treemap_text(fontface = "italic", colour = "white", place = "centre",
                    grow = TRUE)
```



Spesa delle famiglie

Import subset

```
df_spesa = df[df$INDICATORE == "Spesa delle famiglie",]
```

Retrieve NUTS2 code from NUTS3 (deleting the last character)

```
df_spesa$REGION = gsub('.{1}$', '', df_spesa$CODICE.NUTS.3.2021)
```

Substitute NUTS2 cose with region name

```
df_spesa[df_spesa$REGION == "ITC1",]$REGION <- "Piemonte"
df_spesa[df_spesa$REGION == "ITC2",]$REGION <- "Valle d'Aosta/Vallée d'Aoste"
df_spesa[df_spesa$REGION == "ITC3",]$REGION <- "Liguria"
df_spesa[df_spesa$REGION == "ITC4",]$REGION <- "Lombardia"
df_spesa[df_spesa$REGION == "ITH1",]$REGION <- "Trentino-Alto Adige/Südtirol"
```



```
df_spesa[df_spesa$REGION == "ITH2",]$REGION <- "Trentino-Alto Adige/Südtirol"
df_spesa[df_spesa$REGION == "ITH3",]$REGION <- "Veneto"
df_spesa[df_spesa$REGION == "ITH4",]$REGION <- "Friuli-Venezia Giulia"
df_spesa[df_spesa$REGION == "ITH5",]$REGION <- "Emilia-Romagna"
df_spesa[df_spesa$REGION == "ITI1",]$REGION <- "Toscana"
df_spesa[df_spesa$REGION == "ITI2",]$REGION <- "Umbria"
df_spesa[df_spesa$REGION == "ITI3",]$REGION <- "Marche"
df_spesa[df_spesa$REGION == "ITI4",]$REGION <- "Lazio"
df_spesa[df_spesa$REGION == "ITF1",]$REGION <- "Abruzzo"
df_spesa[df_spesa$REGION == "ITF2",]$REGION <- "Molise"
df_spesa[df_spesa$REGION == "ITF3",]$REGION <- "Campania"
df_spesa[df_spesa$REGION == "ITF4",]$REGION <- "Puglia"
df_spesa[df_spesa$REGION == "ITF5",]$REGION <- "Basilicata"
df_spesa[df_spesa$REGION == "ITF6",]$REGION <- "Calabria"
df_spesa[df_spesa$REGION == "ITG1",]$REGION <- "Sicilia"
df_spesa[df_spesa$REGION == "ITG2",]$REGION <- "Sardegna"
```

Calculate the mean of value for each region

```
means <- aggregate(VALORE ~ REGION, data = df_spesa, FUN = mean)
```

Add column with the means to the dataset

```
df_spesa = df_spesa %>%
  right_join(. , means, by=c("REGION"="REGION"))
```

Import geojson for choropleth map graph

```
spdf = geojson_read("/Users/elypa/limits_IT_regions.geojson", what = "sp")
spdf_fortified = tidy(spdf, region = "reg_name")
```

Join the geojson with the dataset

```
spdf_fortified = spdf_fortified %>%
  left_join(. , df_spesa, by=c("id"="REGION"))
```

Plot the choropleth map graph

```
# Plot the result
choropleth_ita_tot_GDP = ggplot() +

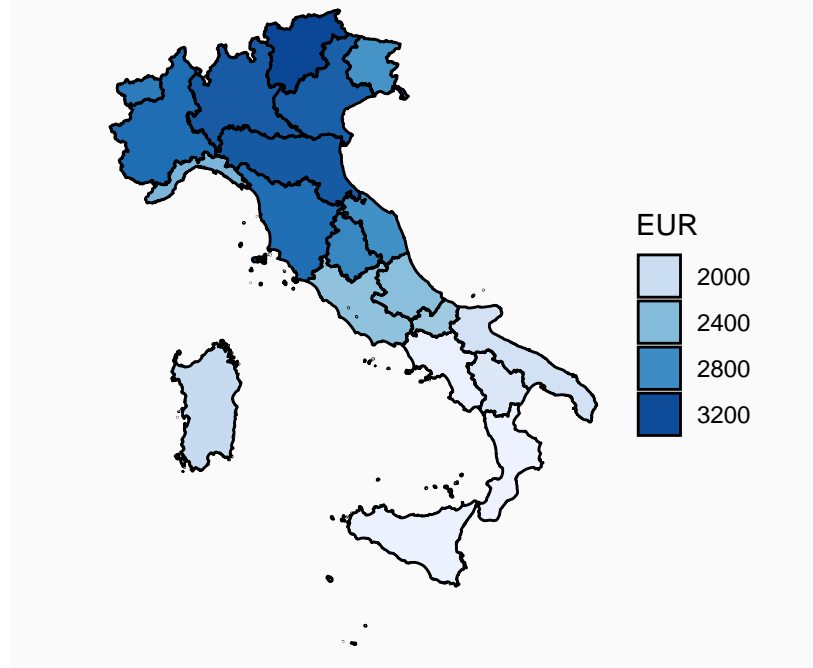
# pass info on the map
  geom_polygon(data = spdf_fortified, aes(fill = `VALORE.y`, x = long, y = lat, group = group), color="black")
```

```
# customize theme and title
labs(title="Spesa delle famiglie",
      subtitle="Per il consumo di beni durevoli - In euro all'anno") +
theme_void() +
theme(
  plot.title = element_text(size= 30, hjust=0.5, color = "black", margin = margin(b = -0.1, t = 0.4, l = 0, r = 0)),
  plot.subtitle = element_text(size= 17, hjust=0.5, color = "black", margin = margin(b = -0.1, t = 0.4, l = 0, r = 0)),
  plot.background = element_rect(fill = "#fafafa", color = NA),
  plot.margin = unit(c(0, 1, 0, 1), "cm")
) +
guides(fill=guide_legend("EUR")) +
coord_map() +
scale_fill_distiller(guide = FALSE, direction = 1, limits=c(1790,3240))

choropleth_ita_tot_GDP
```

Spesa delle famiglie

Per il consumo di beni durevoli – In euro all'anno



Spettacoli - Spesa al botteghino

Import subset

```
df_spett = df[df$INDICATORE == "Spettacoli - Spesa al botteghino",]
```

Retrieve NUTS2 code from NUTS3 (deleting the last character)

```
df_spett$REGION = gsub('.{1}$', '', df_spett$CODICE.NUTS.3.2021)
```

Substitute NUTS2 code with region name

```
df_spett[df_spett$REGION == "ITC1",]$REGION <- "Piemonte"
df_spett[df_spett$REGION == "ITC2",]$REGION <- "Valle d'Aosta/Vallée d'Aoste"
df_spett[df_spett$REGION == "ITC3",]$REGION <- "Liguria"
df_spett[df_spett$REGION == "ITC4",]$REGION <- "Lombardia"
df_spett[df_spett$REGION == "ITH1",]$REGION <- "Trentino-Alto Adige/Südtirol"
df_spett[df_spett$REGION == "ITH2",]$REGION <- "Trentino-Alto Adige/Südtirol"
df_spett[df_spett$REGION == "ITH3",]$REGION <- "Veneto"
df_spett[df_spett$REGION == "ITH4",]$REGION <- "Friuli-Venezia Giulia"
df_spett[df_spett$REGION == "ITH5",]$REGION <- "Emilia-Romagna"
df_spett[df_spett$REGION == "ITI1",]$REGION <- "Toscana"
df_spett[df_spett$REGION == "ITI2",]$REGION <- "Umbria"
df_spett[df_spett$REGION == "ITI3",]$REGION <- "Marche"
df_spett[df_spett$REGION == "ITI4",]$REGION <- "Lazio"
df_spett[df_spett$REGION == "ITF1",]$REGION <- "Abruzzo"
df_spett[df_spett$REGION == "ITF2",]$REGION <- "Molise"
df_spett[df_spett$REGION == "ITF3",]$REGION <- "Campania"
df_spett[df_spett$REGION == "ITF4",]$REGION <- "Puglia"
df_spett[df_spett$REGION == "ITF5",]$REGION <- "Basilicata"
df_spett[df_spett$REGION == "ITF6",]$REGION <- "Calabria"
df_spett[df_spett$REGION == "ITG1",]$REGION <- "Sicilia"
df_spett[df_spett$REGION == "ITG2",]$REGION <- "Sardegna"
```

Calculate the mean of value for each region

```
means <- aggregate(VALORE ~ REGION, data = df_spett, FUN = mean)
```

Barplot

```
ggplot(means, aes(x=reorder(REGION, VALORE), y=VALORE)) +
  geom_bar(stat = "identity", fill= "#17a2b8") +
  geom_hline(yintercept = mean(means$VALORE),
            color = "tomato", size=1) +
  coord_flip() +
  xlab("") +
  ylab("Euro pro capite all'anno")
```

