

# Colombia COVID-19 - Central region

---

Angela Carraro, Giullia Monteiro Milano Oliveira, Gaia Saveri  
3/07/2020

University of Trieste

We decided to do **central Colombia**, basically because it is where the capital is.

We built a model for the number of **confirmed** cases using all the others covariates (plus some we created) and we estimated the predictive accuracy of our selected model.

# Preprocessing

```
colombia_covid <- as.data.frame(read_csv("data/datasets_567855_1056808_Casos1.csv"))
colnames(colombia_covid)[5] <- "Atención"
colnames(colombia_covid)[8] <- "Tipo"
#slice the main dataset
central.colombia.dep <- c("Bogotá D.C.", "Tolima", "Cundinamarca", "Meta", "Boyacá", "Valle del Cauca", "Risaralda", "Caldas", "Boyacá", "Antioquia", "Santander")
central.colombia.rows <- which(colombia_covid$`Departamento o Distrito` %in% central.colombia.dep)
colombia_covid <- colombia_covid[central.colombia.rows, ]

head(colombia_covid)
```

```
##   ID de caso Fecha de diagnóstico Ciudad de ubicación Departamento o Distrito
## 1         1      2020-03-06          Bogotá          Bogotá D.C.
## 2         2      2020-03-09          Buga          Valle del Cauca
## 3         3      2020-03-09        Medellín          Antioquia
## 4         4      2020-03-11        Medellín          Antioquia
## 5         5      2020-03-11        Medellín          Antioquia
## 6         6      2020-03-11          Itagüí          Antioquia
##   Atención Edad Sexo      Tipo País de procedencia Grupo de edad
## 1 Recuperado  19   F  Importado          Italia      19-30
## 2 Recuperado  34   M  Importado          España      31-45
## 3 Recuperado  50   F  Importado          España      46-60
## 4 Recuperado  55   M Relacionado          Colombia      46-60
```