

Case Study 3: Visualization

Nicolas Bschor

2022-05-25

```
colors <- c("#006699", "#E18922")

col_names <- c("country_name", "official_name_en", "iso_3166_2", "iso_3166_3",
               "region_name", "sub_region_name", "developed", "age_median",
               "youth_unempl_rate", "above_average_median_age", "above_average_yu")

df_vars <- read.csv("df_vars.csv", sep = ";", header = FALSE, col.names=col_names, na.strings = ".", de

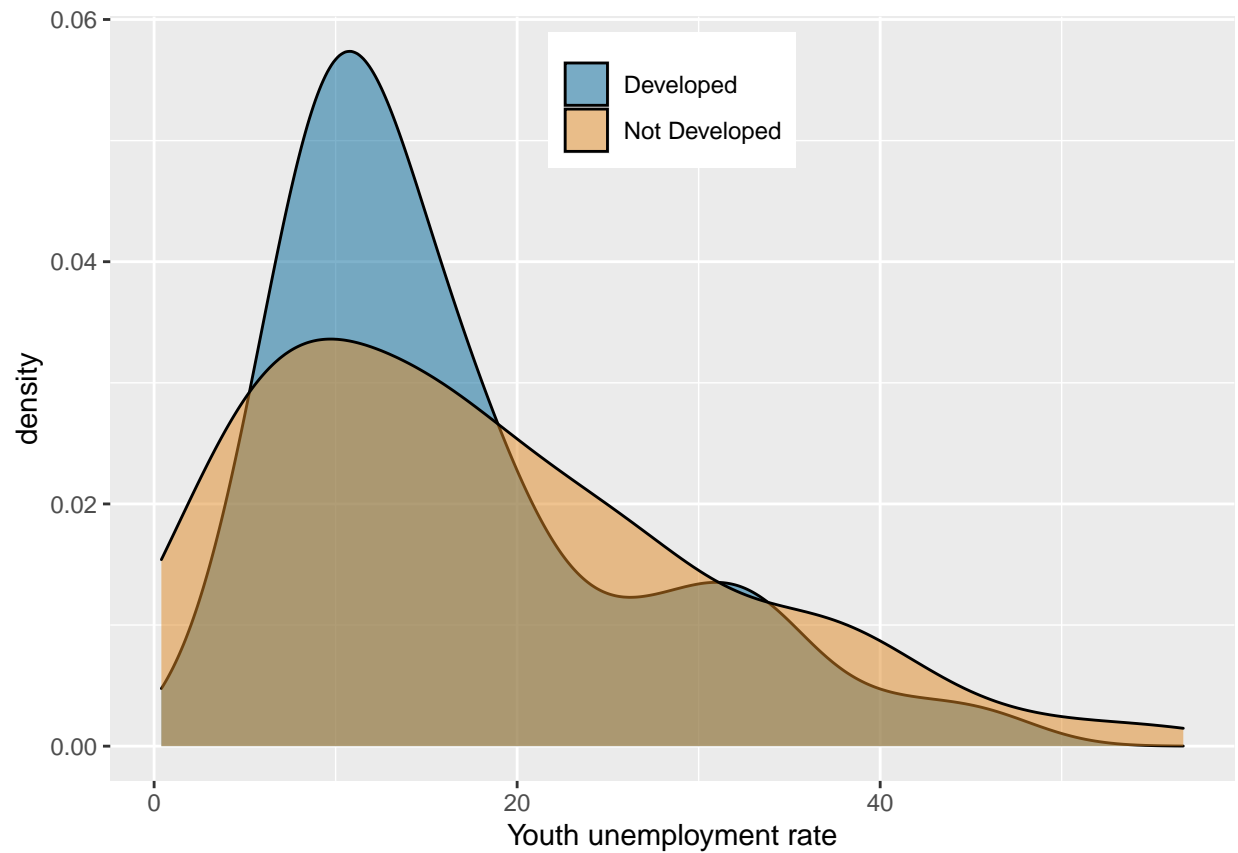
df_vars <- df_vars %>%
  filter(!is.na(`developed`) & !is.na(`iso_3166_2`)) %>%
  mutate(`above_average_median_age` = ifelse(is.na(`above_average_median_age`), NA, `above_average_medi
  mutate(`above_average_yu` = ifelse(is.na(`above_average_yu`), NA, `above_average_yu`=='yes'))

df_vars %>%
  mutate(development = ifelse(`developed`, 'Developed', 'Not Developed')) %>%
  ggplot(aes(x = `age_median`, group = `development`)) +
  geom_density(aes(fill = `development`, alpha = 0.5, color="black")) +
  xlab("Median age of population") +
  guides(fill = guide_legend(title = NULL)) +
  theme(legend.justification = c("center", "top"), legend.position = c(0.5, 0.98))+
  scale_fill_manual("legend", values = c("Developed" = "#006699", "Not Developed" = "#E18922"))
```

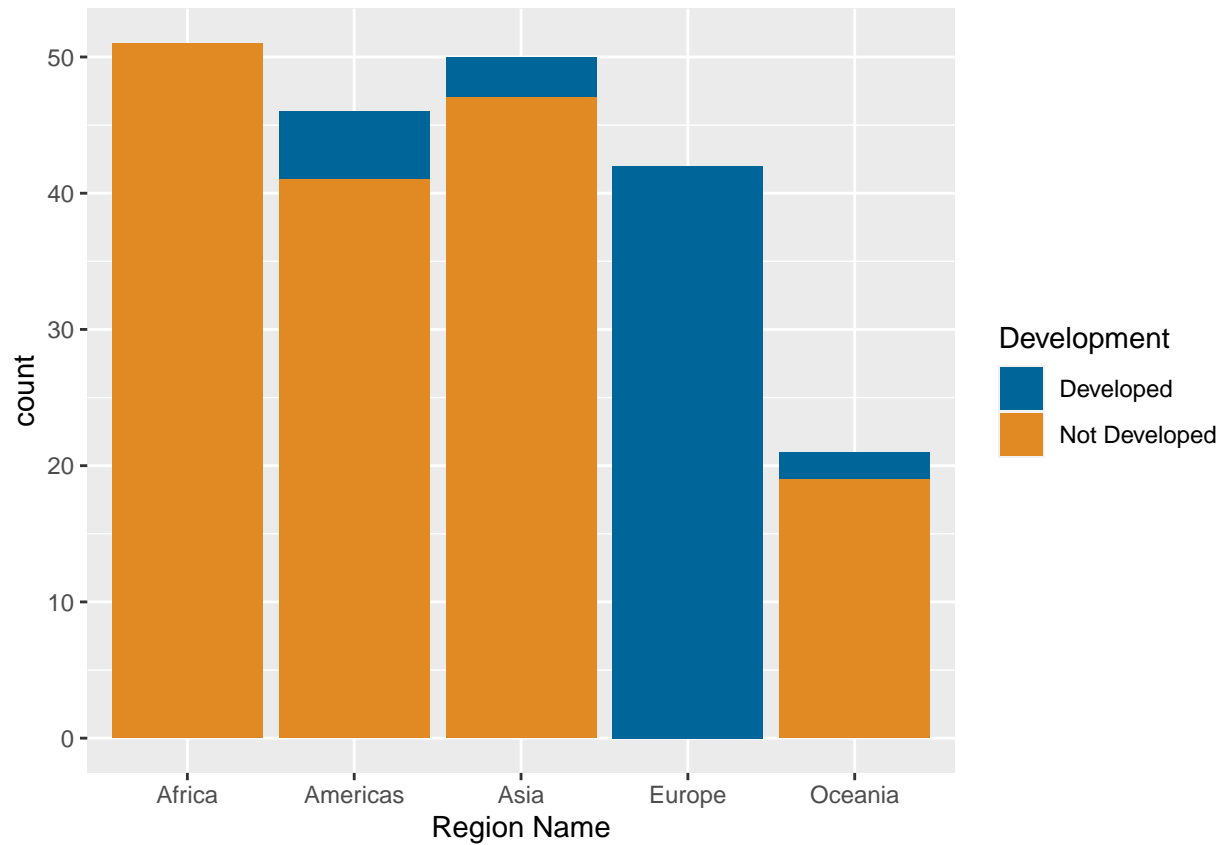


```
df_vars %>%
  mutate(development = ifelse(`developed`, 'Developed', 'Not Developed')) %>%
  ggplot(aes(x = `youth_unempl_rate`, group = `development`)) +
  geom_density(aes(fill = `development`), alpha = 0.5, color="black") +
  xlab("Youth unemployment rate") +
  guides(fill = guide_legend(title = NULL)) +
  theme(legend.justification = c("center", "top"), legend.position = c(0.5, 0.98)) +
  scale_fill_manual("legend", values = c("Developed" = "#006699", "Not Developed" = "#E18922"))
```

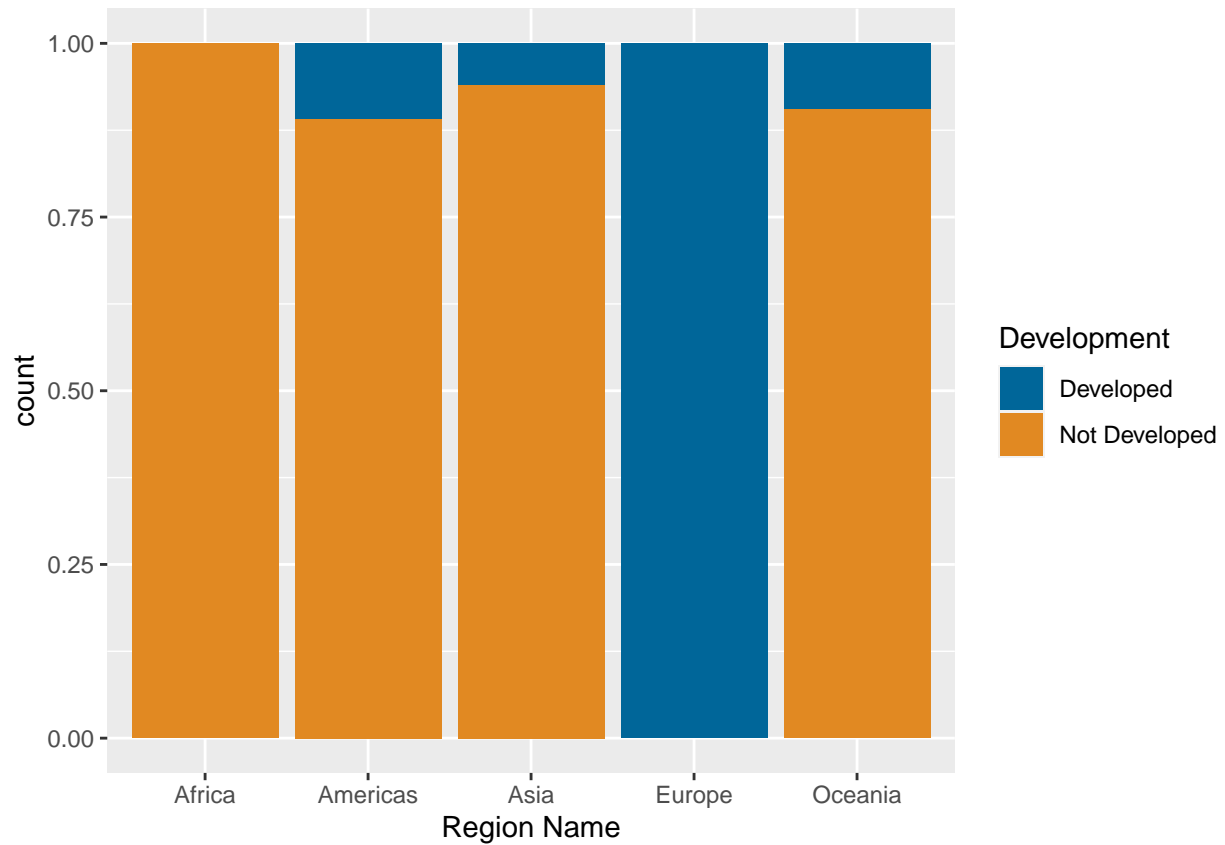
```
## Warning: Removed 39 rows containing non-finite values (stat_density).
```



```
df_vars %>%
  mutate(Development = ifelse(`developed`, 'Developed', 'Not Developed')) %>%
  ggplot(aes(x = `region_name`, fill = `Development`)) +
  geom_bar() +
  xlab("Region Name") +
  scale_fill_manual("Development", values = c("Developed" = "#006699", "Not Developed" = "#E18922"))
```



```
df_vars %>%
  mutate(Development = ifelse(`developed`, 'Developed', 'Not Developed')) %>%
  ggplot(aes(x = `region_name`, fill = `Development`)) +
  geom_bar(position = "fill") +
  xlab("Region Name") +
  scale_fill_manual("Development", values = c("Developed" = "#006699", "Not Developed" = "#E18922"))
```

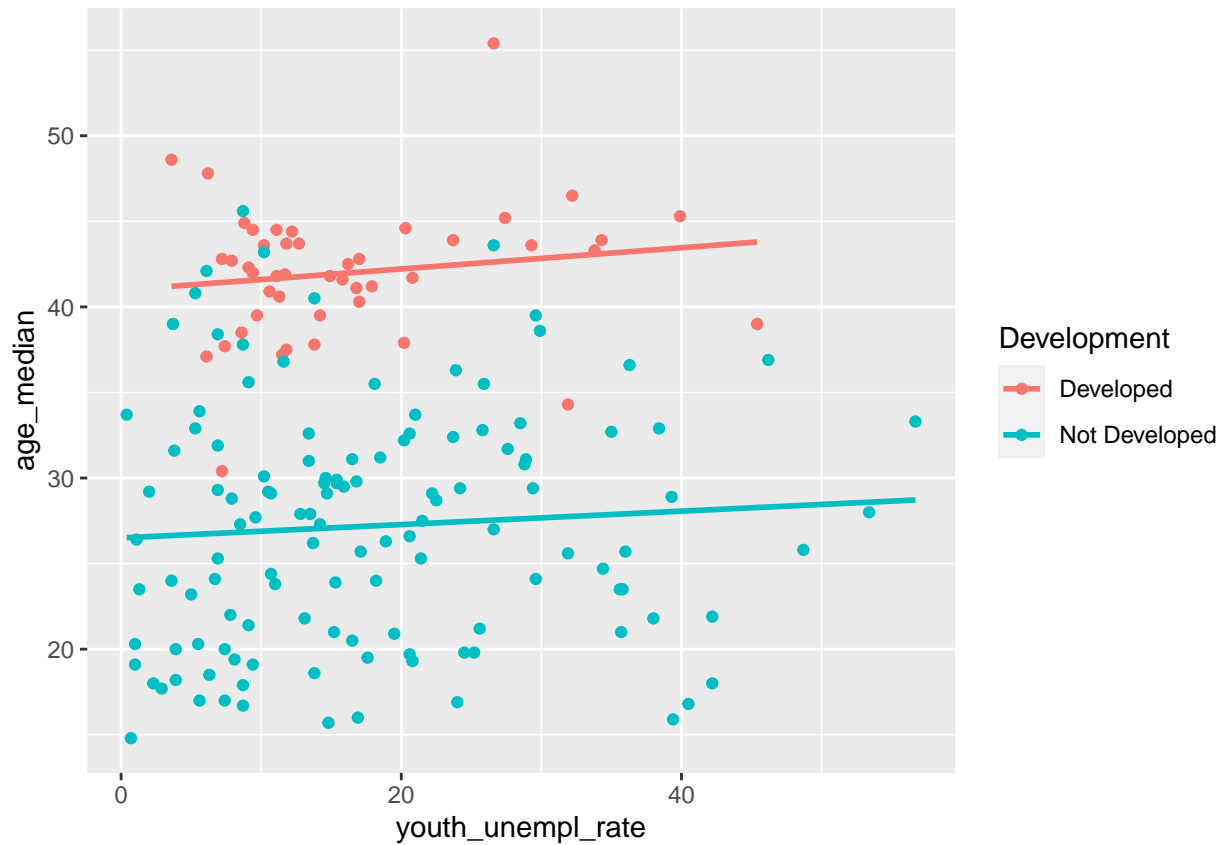


```
df_vars %>%
  mutate(Development = ifelse(`developed`, 'Developed', 'Not Developed')) %>%
  ggplot(aes(x=`youth_unempl_rate`, y=`age_median`, color=`Development`)) +
  geom_point() +
  geom_smooth(method='lm', se=FALSE)
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

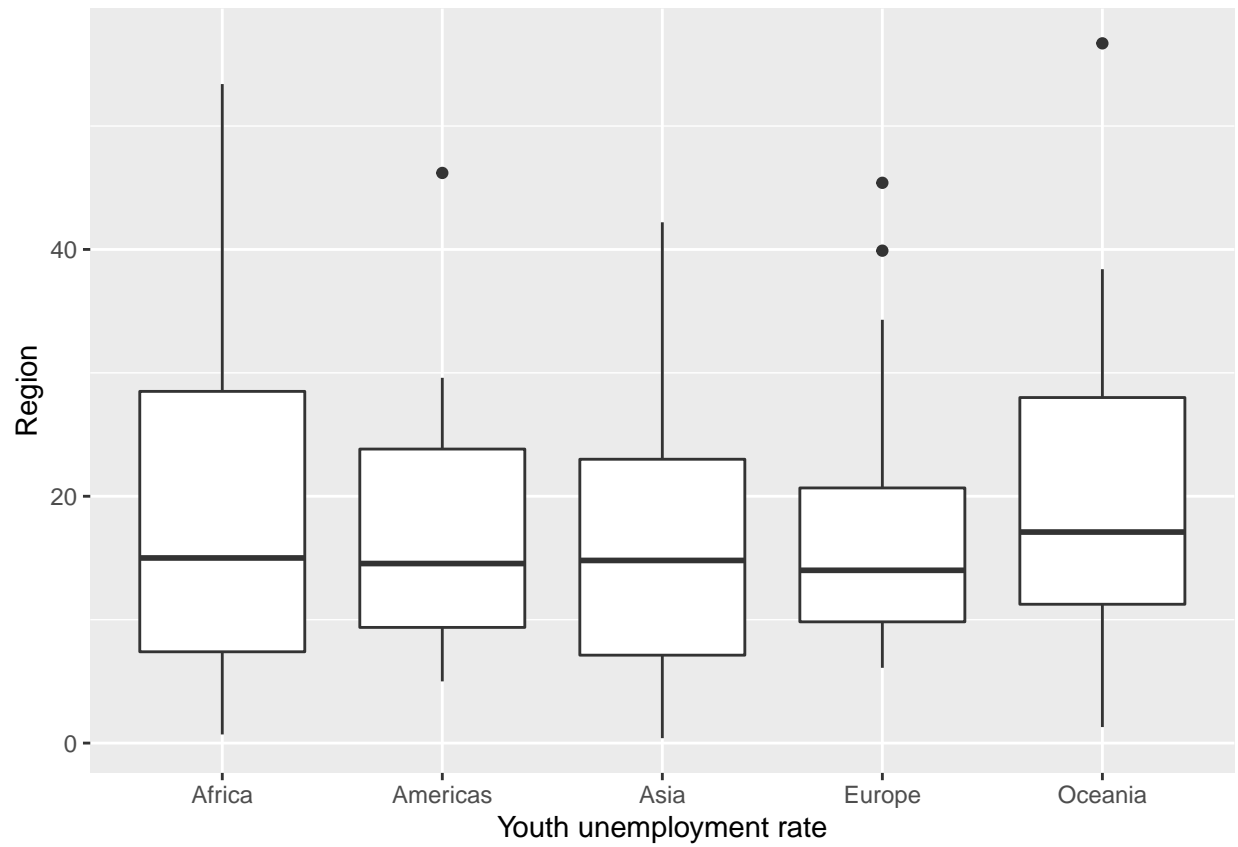
```
## Warning: Removed 39 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 39 rows containing missing values (geom_point).
```

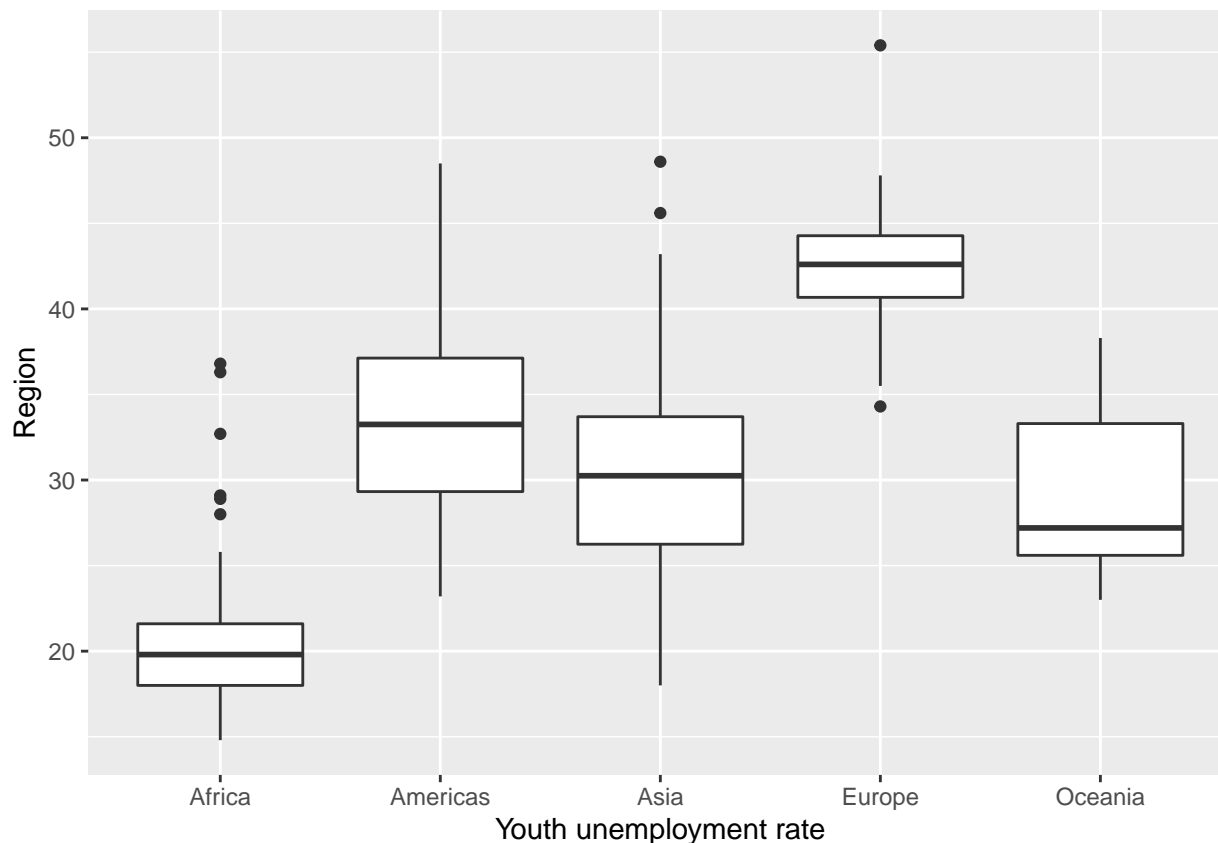


```
df_vars %>%
  ggplot(aes(x=`region_name`, y=`youth_unempl_rate`)) +
  geom_boxplot( ) +
  xlab("Youth unemployment rate") +
  ylab("Region")
```

```
## Warning: Removed 39 rows containing non-finite values (stat_boxplot).
```



```
df_vars %>%  
  ggplot(aes(x=`region_name`, y=`age_median`)) +  
  geom_boxplot( ) +  
  xlab("Youth unemployment rate") +  
  ylab("Region")
```



```
cbPalette <- c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")
```

```
df_vars %>%
  group_by(`region_name`, `sub_region_name`) %>%
  summarize(avg_yur = mean(`youth_unempl_rate`, na.rm=TRUE)) %>%

  ggplot(aes(x=`avg_yur`, y=fct_reorder(`sub_region_name`, `avg_yur`), color=`region_name`)) +
  geom_point(size=4) +
  xlab("Average youth unemployment rate") +
  ylab("Sub-Region") +
  scale_colour_discrete("Continents") +
  scale_color_manual(values = cbPalette)
```

```
## 'summarise()' has grouped output by 'region_name'. You can override using the
## '.groups' argument.
## Scale for 'colour' is already present. Adding another scale for 'colour', which
## will replace the existing scale.
```