

Case Study 3

AKSTA Statistical Computing

Fani Sentinella-Jerbić

30.03.2023

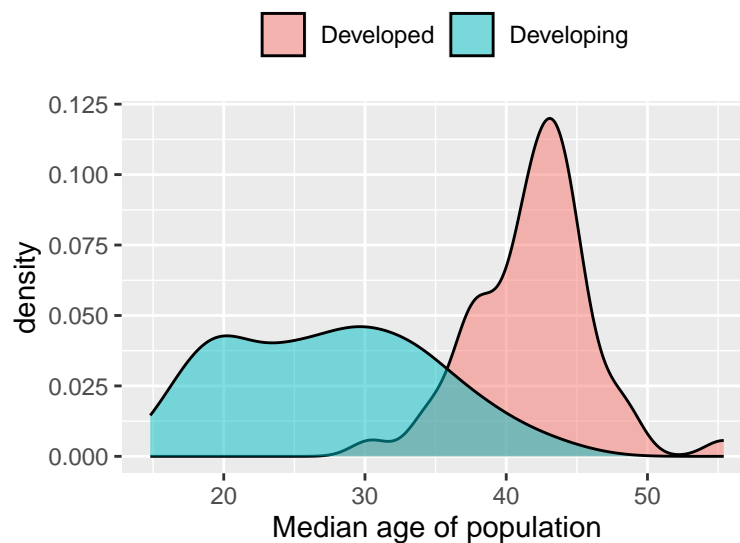
Note: I'm ignoring Taiwan based on my previous assignment.

1.

Density plots of the **median age** in developing and developed countries:

```
library(dplyr)
library(ggplot2)
data <- read.csv("final_dataset.csv", sep=";", ) %>%
  na_if('.') %>%
  filter(!(country=='Taiwan'))
data$youth_unempl_rate = as.numeric(as.character(data$youth_unempl_rate))
```

```
ggplot(data) +
  geom_density(aes(x=median_age, fill=dev), alpha=0.5) +
  theme(legend.title=element_blank(), legend.position="top") +
  xlab("Median age of population")
```

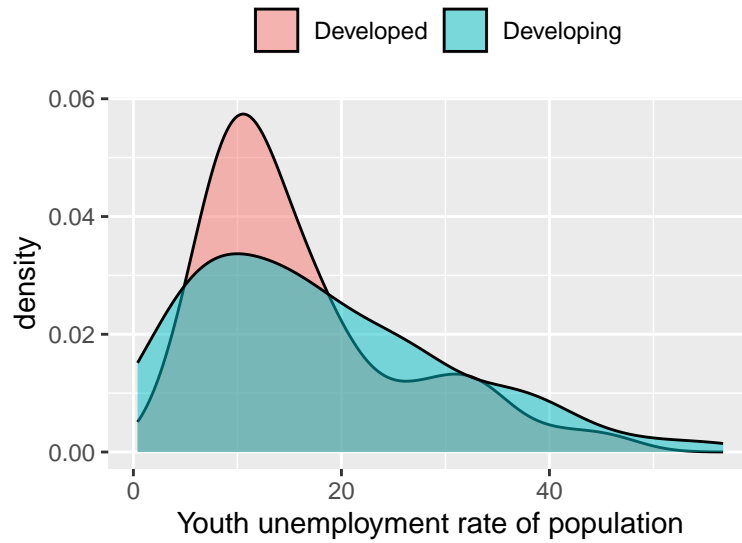


Population in developed countries is older and less

2.

Density plots of the **youth unemployment rate** in developing and developed countries:

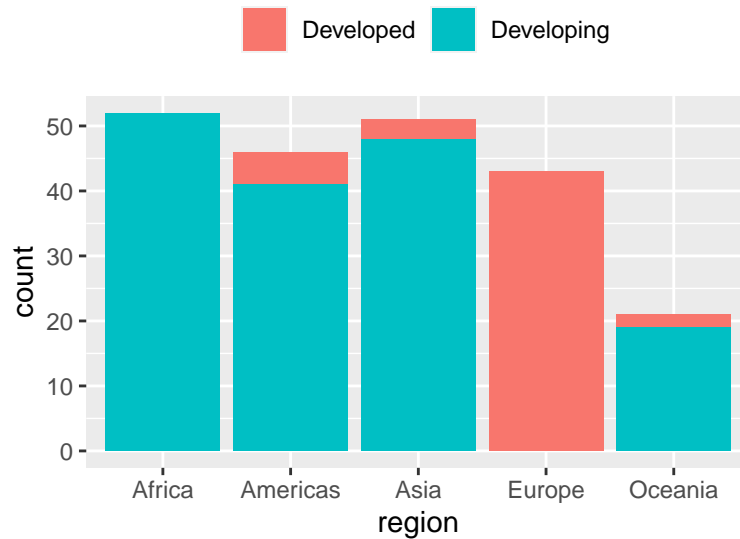
```
ggplot(data) +  
  geom_density(aes(x=youth_unempl_rate, fill=dev), alpha=0.5) +  
  theme(legend.title=element_blank(), legend.position="top") +  
  xlab("Youth unemployment rate of population")
```



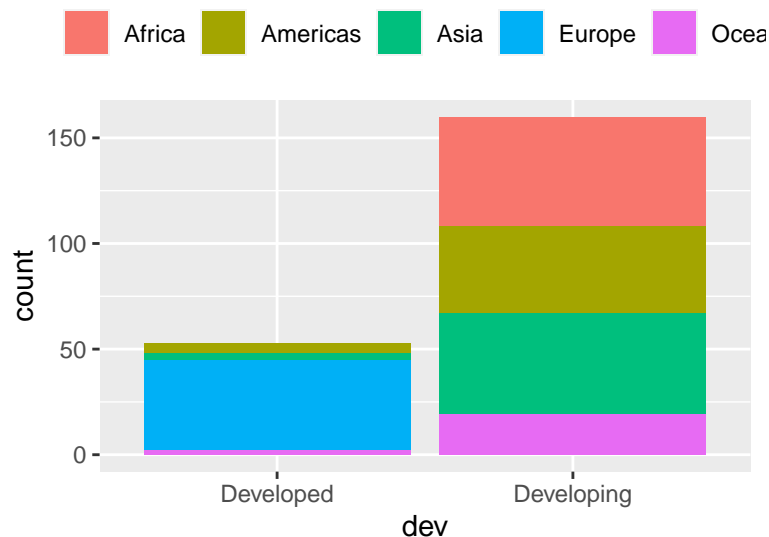
3.

Stacked barplots of **absolute frequencies** showing how the entities are split into regions and development status: (Note: because of ambiguity in assignment text I'm plotting both ways.)

```
data %>% group_by(region)%>%
  ggplot(aes(fill=dev, x=region)) +
  geom_bar(position="stack") +
  theme(legend.title=element_blank(),legend.position="top")
```

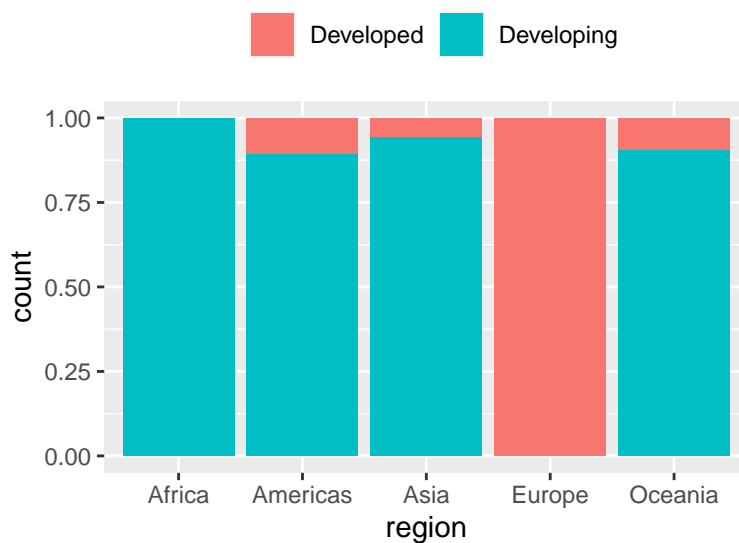


```
data %>% group_by(region)%>%
  ggplot(aes(fill=region, x=dev)) +
  geom_bar(position="stack") +
  theme(legend.title=element_blank(),legend.position="top")
```

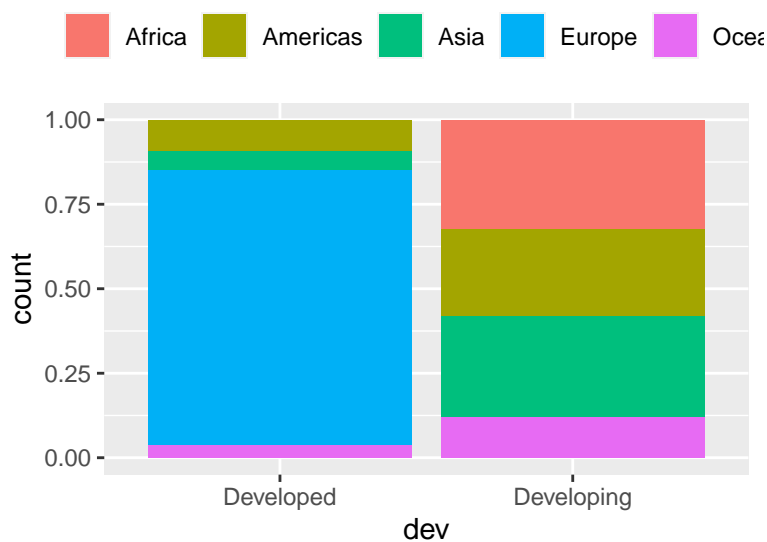


Stacked barplots of **relative frequencies** showing how the entities are split into regions and development status:

```
data %>% group_by(region)%>%
  ggplot(aes(fill=dev, x=region)) +
  geom_bar(position="fill") +
  theme(legend.title=element_blank(),legend.position="top")
```



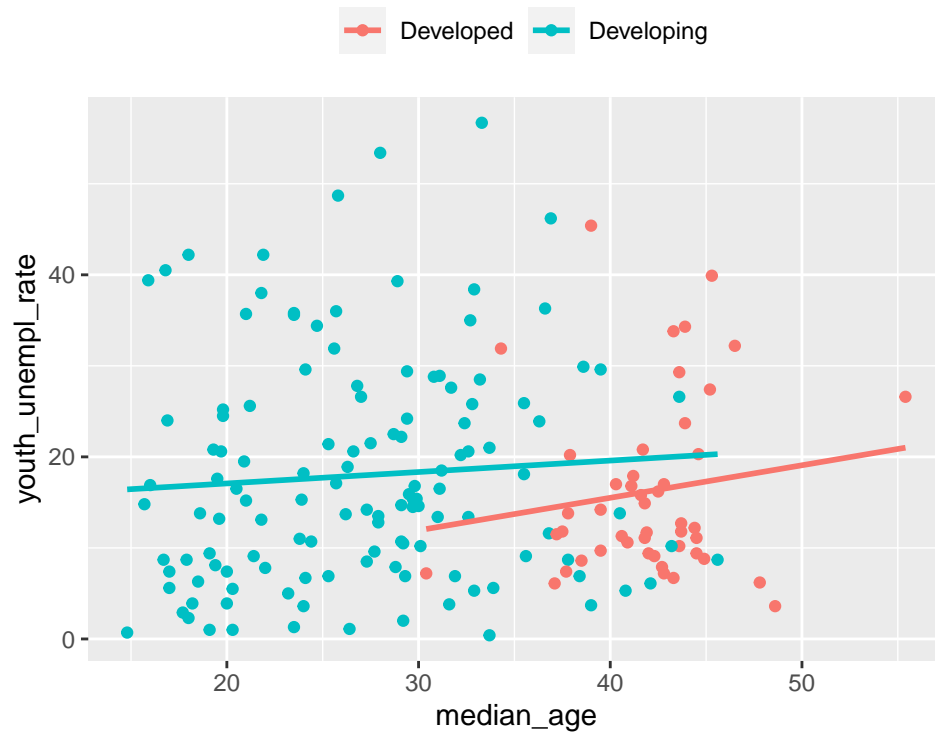
```
data %>% group_by(region)%>%
  ggplot(aes(fill=region, x=dev)) +
  geom_bar(position="fill") +
  theme(legend.title=element_blank(),legend.position="top")
```



4.

Relationship between median age and youth unemployment rate:

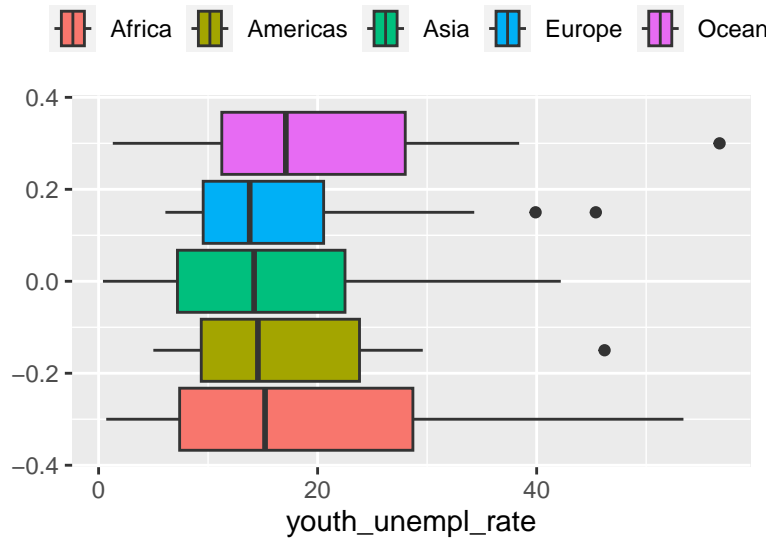
```
ggplot(data, aes(x=median_age, y=youth_unempl_rate, color=dev)) +  
  geom_point() +  
  geom_smooth(method="lm", se=FALSE) +  
  theme(legend.title=element_blank(), legend.position="top")
```



5.

Parallel boxplots of the youth unemployment rate for each region:

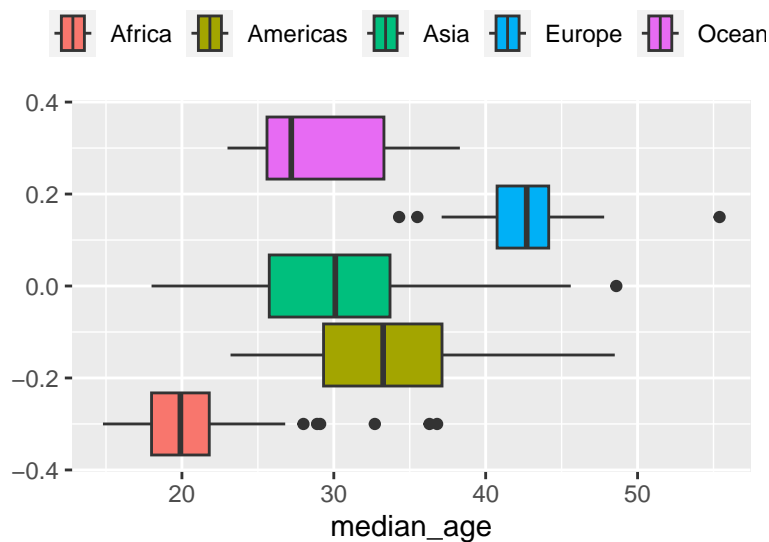
```
ggplot(data, aes(x=youth_unempl_rate, fill=region)) +  
  geom_boxplot() +  
  theme(legend.title=element_blank(), legend.position="top")
```



6.

Parallel boxplots of the median age for each region:

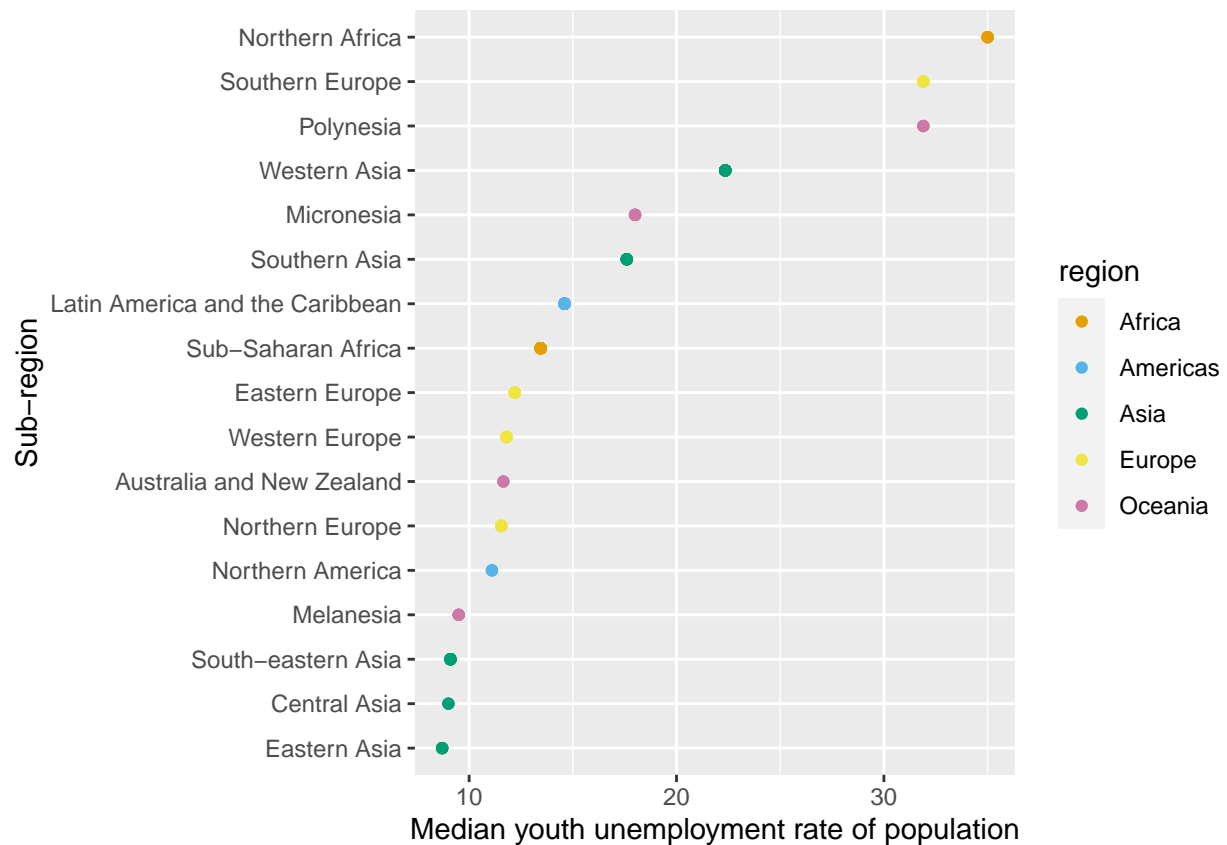
```
ggplot(data, aes(x=median_age, fill=region)) +  
  geom_boxplot() +  
  theme(legend.title=element_blank(), legend.position="top")
```



7.

Median youth unemployment rate per subregion:

```
library(forcats)
cbbPalette <- c("#E69F00", "#56B4E9", "#009E73", "#F0E442", "#CC79A7")
data %>%
  filter(!is.na(youth_unempl_rate)) %>%
  group_by(subregion) %>%
  summarize(medYUR = median(youth_unempl_rate), region=region) %>%
  ggplot(aes(x=medYUR, y=fct_reorder(subregion, medYUR), color=region)) +
  geom_point() +
  scale_colour_manual(values=cbbPalette) +
  xlab("Median youth unemployment rate of population") +
  ylab("Sub-region")
```



8.

9.