

Atividade 1

A seguir a base gapminder muito usada em diversos exemplos de machine learning. Iremos focar na análise da relação entre expectativa de vida (lifeExp) e pib per capita (gdpPercap) para os diversos países.

```
library(modelr)
library(tidyverse)
library(gapminder)

head(gapminder)

## # A tibble: 6 x 6
##   country    continent  year lifeExp      pop gdpPercap
##   <fct>      <fct>    <int>  <dbl>    <int>    <dbl>
## 1 Afghanistan Asia      1952   28.8  8425333    779.
## 2 Afghanistan Asia      1957   30.3  9240934    821.
## 3 Afghanistan Asia      1962   32.0 10267083    853.
## 4 Afghanistan Asia      1967   34.0 11537966    836.
## 5 Afghanistan Asia      1972   36.1 13079460    740.
## 6 Afghanistan Asia      1977   38.4 14880372    786.

dim(gapminder)

## [1] 1704    6

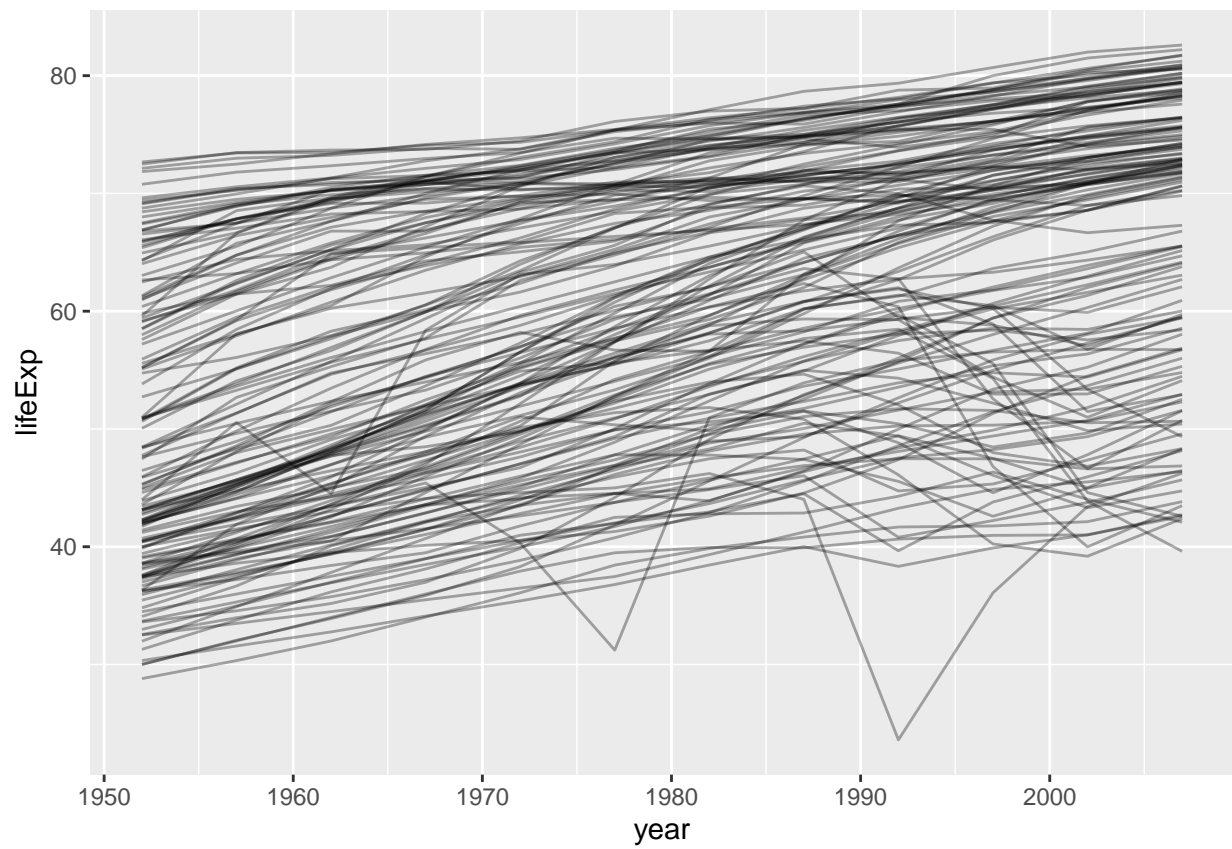
nlevels(gapminder$country) ### 142

## [1] 142

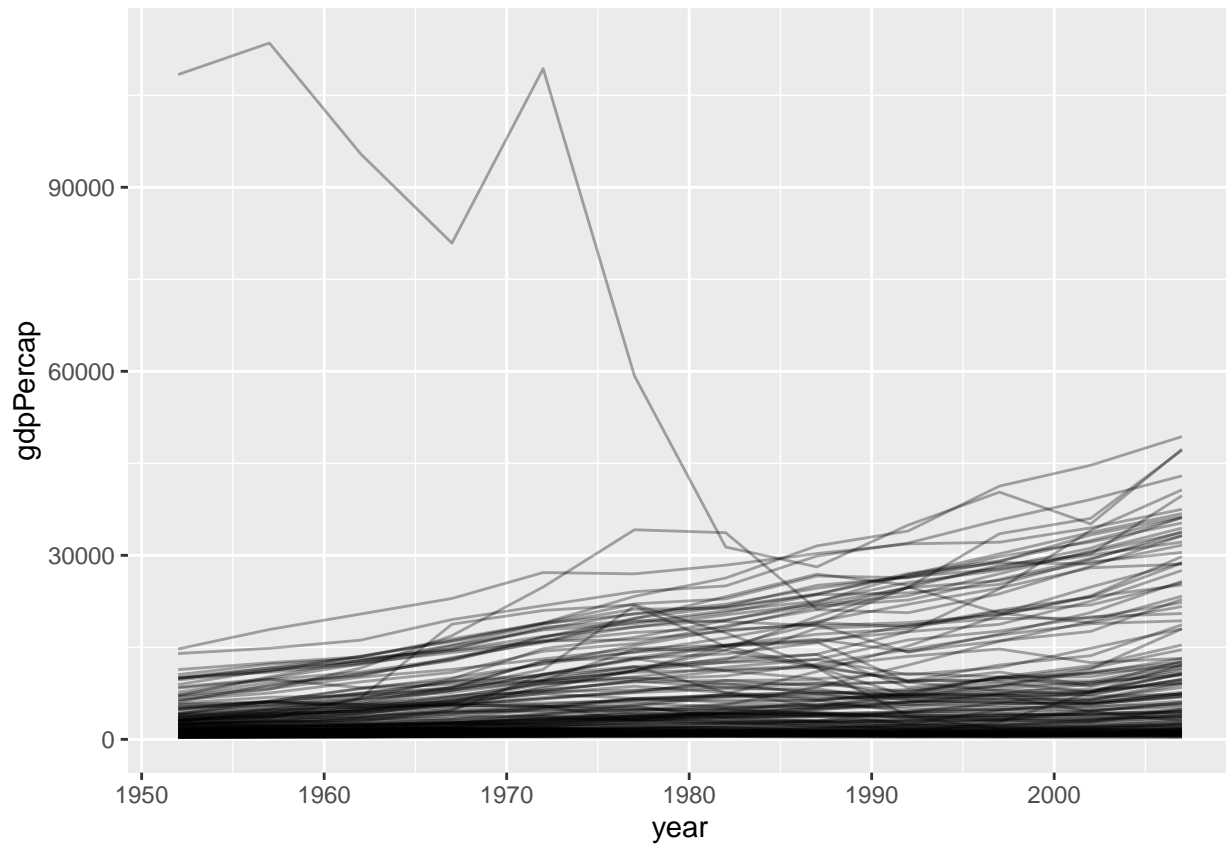
length(unique(gapminder$year)) ### 12

## [1] 12

gapminder %>%
  ggplot(aes(year, lifeExp, group = country)) +
  geom_line(alpha = 1/3)
```



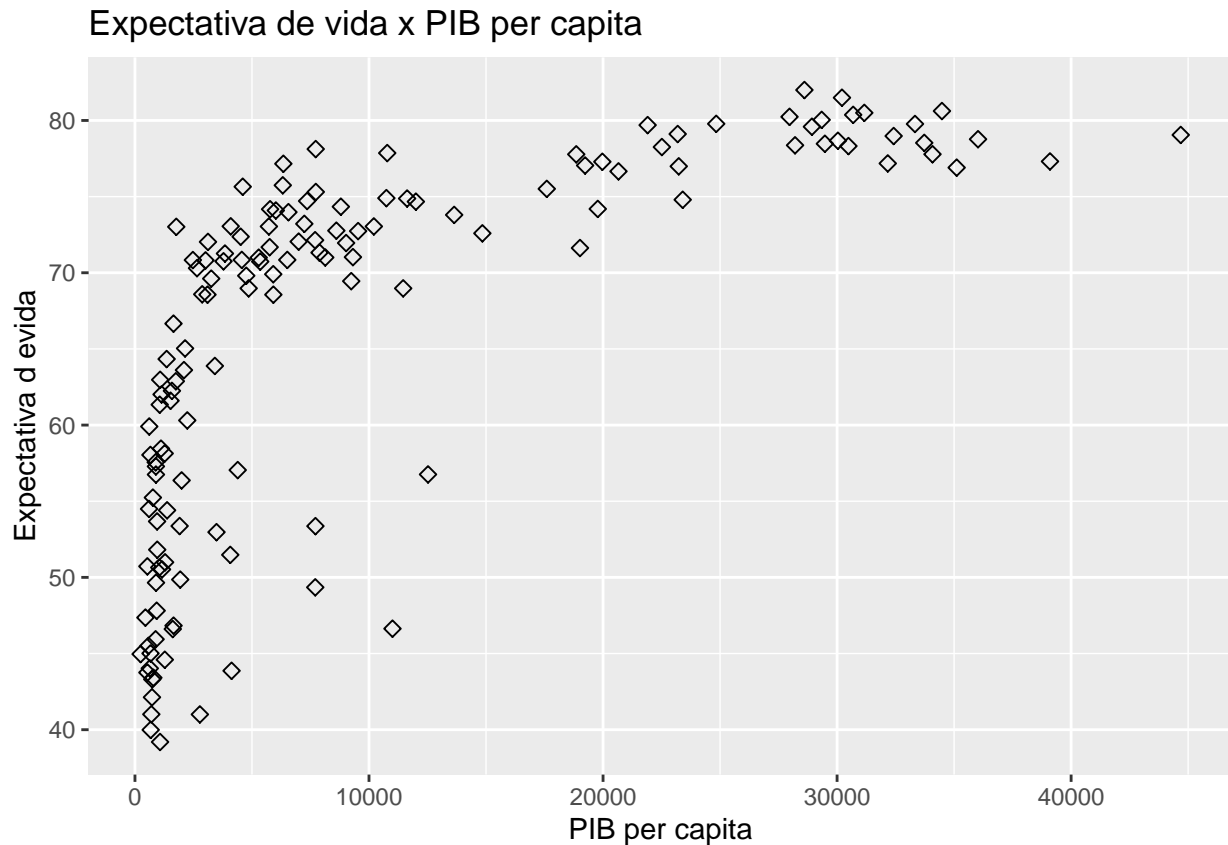
```
gapminder %>%  
  ggplot(aes(year, gdpPercap, group = country)) +  
  geom_line(alpha = 1/3)
```



Em particular iremos selecionar o ano de 2002 para essa análise.

```
gap.dt <- filter(gapminder, year == "2002")

ggplot(gap.dt, aes(x=gdpPercap, y=lifeExp)) +
  geom_point(size=2, shape=23) +
  labs(title="Expectativa de vida x PIB per capita",
        x="PIB per capita", y = "Expectativa d evida")
```



Roteiro para atividade:

- A relação entre x e y parece linear? Considere a transformação \log para x e veja se houve mudança na relação;
- Divida os dados em treinamento e teste e ajuste os seguintes modelos:
 - Modelo 1: dados x e y na escala original;
 - Modelo 2: $\log(x)$ como covariável, y na escala original;
- Construa o scatter plot para os dois ajustes e comente;
- Obtenha a previsão no conjunto de teste para os 2 modelos e calcule medidas de acurácia para comparar os modelos. Comente.
- Para o melhor modelo, verifique se existem países com comportamentos que destoem dos demais. Construa o scatter utilizando os nomes dos países. Retire esses países da análise e refaça o ajuste. Recalcule as medidas de acurácia. O R^2 se alterou? Comente.