ETUDE ANOVA

RÉALISÉ PAR : BENABOU MOHAMED EL GHALI

ENCADRÉ PAR : PR. AMAR AMINE

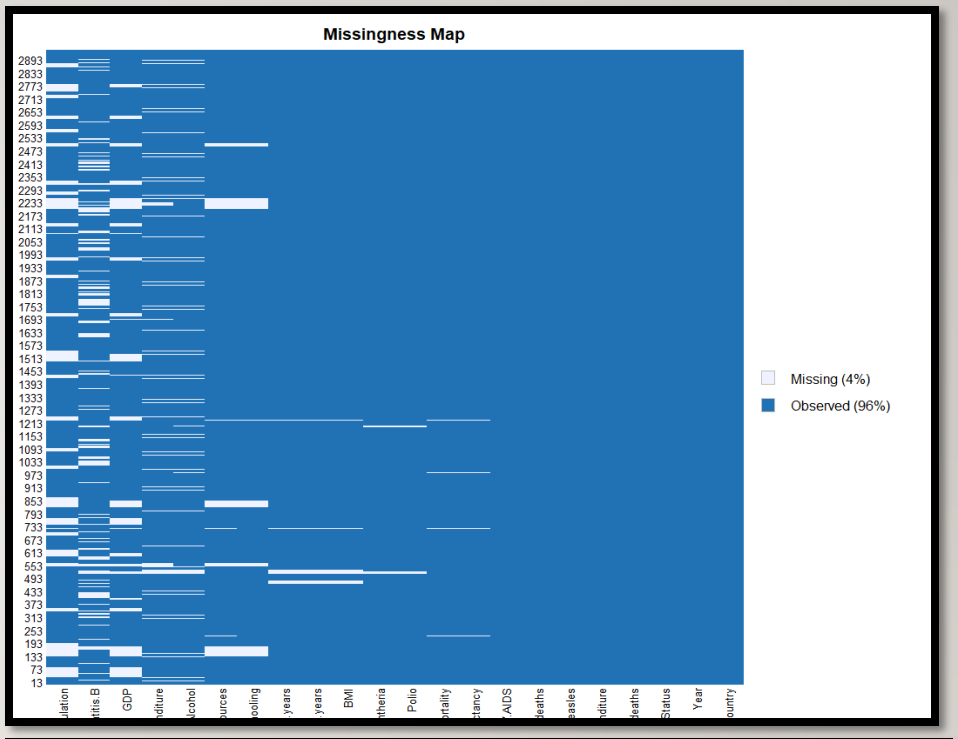
PLAN DE TRAVAIL

- Motivation
- Description du plan d'expérimentation
- Validation des hypothèses d'application
- ANOVA

MOTIVATION

PLAN D'EXPÉRIMENTATION

- Inspection visuelle (MissMap)
- Echantillonnage de taille 32 (pays)
- Variables : Year, Status et Life.expectancy



Year	Status	Life.expectancy
2015	Developed	82.8
2010	Developed	81.9
2005	Developed	81.0
2000	Developed	79.5

```
> summary(data_2000$Life.expectancy)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  39.00   65.70   73.30   70.11   77.80   81.10
> summary(data_2005$Life.expectancy)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  43.30   68.20   74.05   71.92   78.83   88.00
> summary(data_2010$Life.expectancy)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  48.10   69.00   74.35   73.85   81.00   89.00
> summary(data_2015$Life.expectancy)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  51.00   69.78   75.90   74.92   81.62   88.00
>
> summary(data_developed$Life.expectancy)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  71.00   76.67   79.15   79.06   81.72   89.00
> summary(data_developing$Life.expectancy)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  39.00   61.80   68.30   66.33   73.03   85.00
```

VALIDATION DES HYPOTHÈSES D'APPLICATION

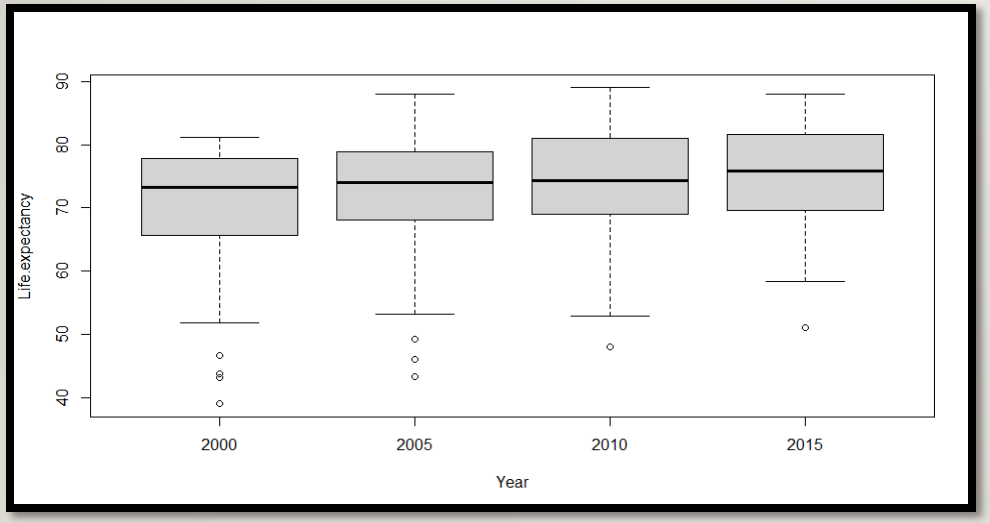
PRÉPARATION DES DONNÉES

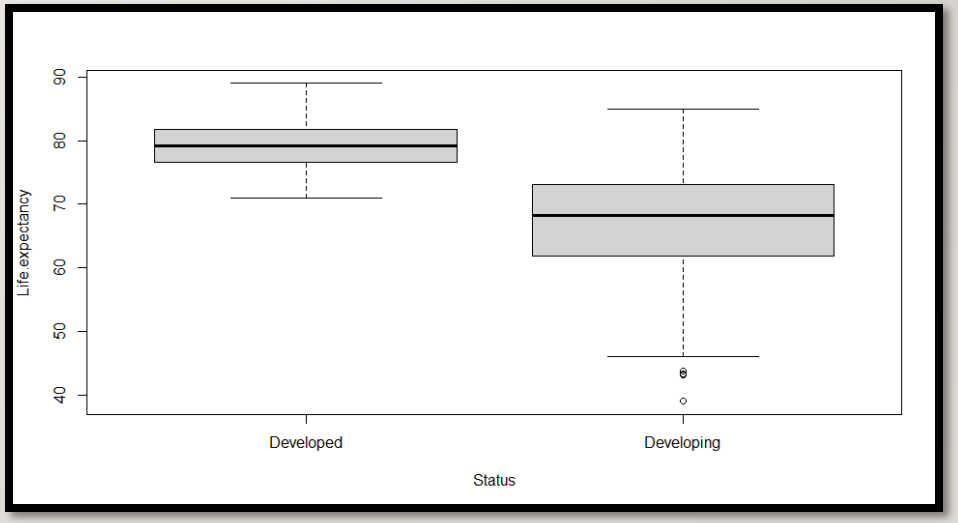
```
library(dplyr)

data = read.csv("C:/Users/ben-g/Downloads/Life Expectancy Data.csv")

data = data %>% select(1: 4)

data = data[data$Year %in% c(2000, 2005, 2010, 2015),]


data_developed = data[data$Status=="Developed",]
view(data_developed)


set.seed(935)
developing = unique(data[data$Status=="Developing",]$Country)

sample = sample(developing, 32)

data_developing = data[data$Country %in% sample,]

data = rbind(data_developed,data_developing)

data = data %>% select(2:4)
```

VALIDATION DES HYPOTHÈSES D'APPLICATION

- **Normalité de la distribution: Etude graphique**

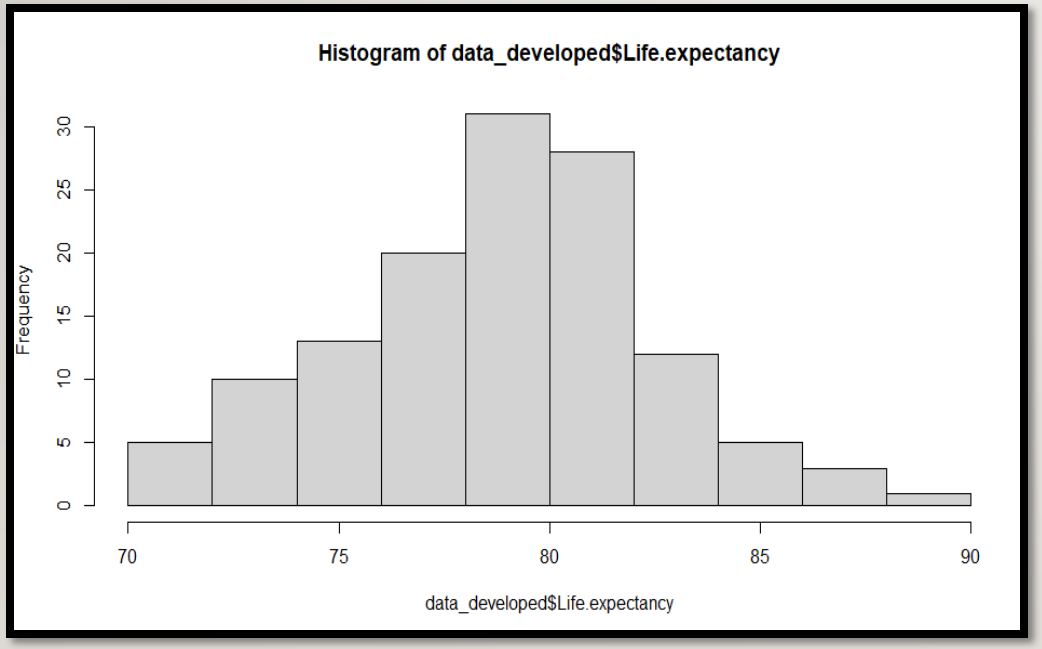
Boîtes à moustaches avec des distributions symétriques pour les variables Status et Year (2005 et 2010)

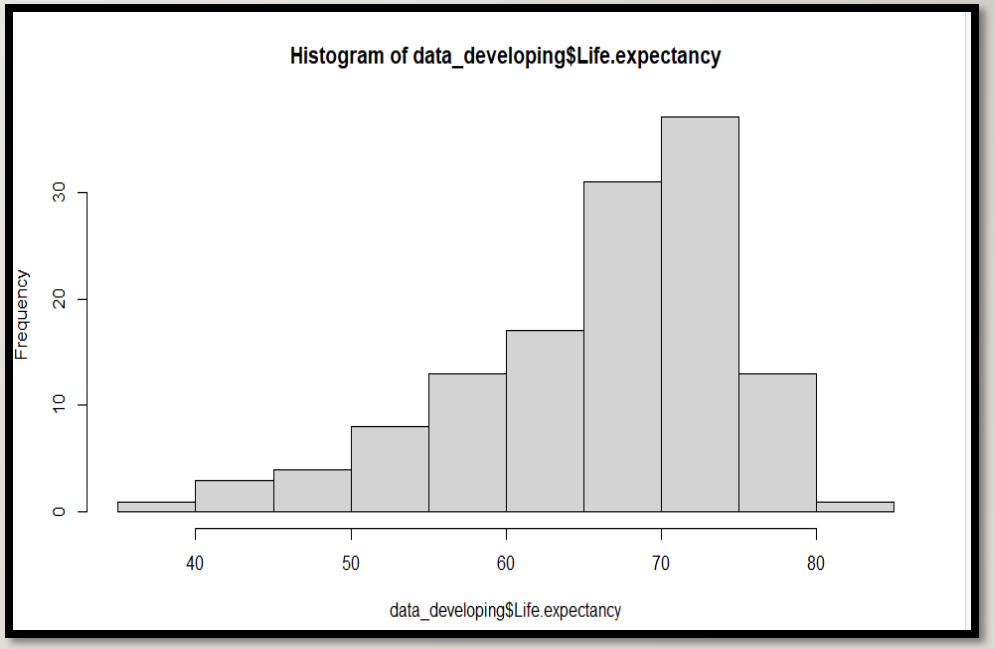




VALIDATION DES HYPOTHÈSES D'APPLICATION

- **Normalité de la distribution: Etude graphique** Histogrammes pour la variable Status

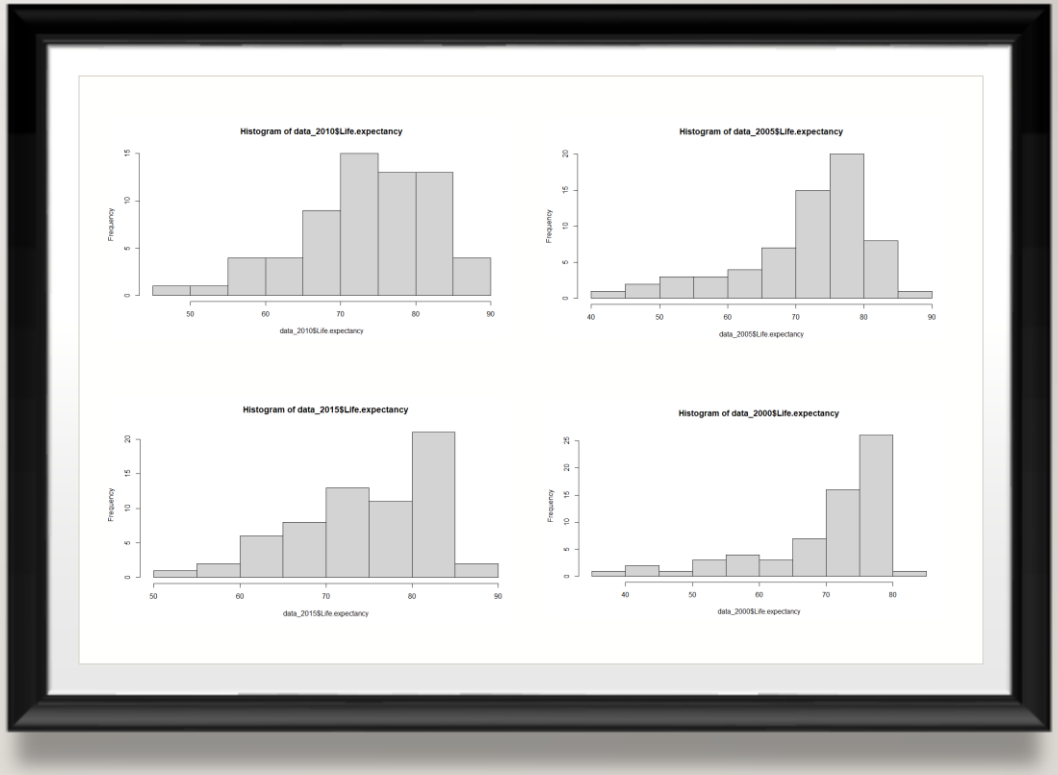




VALIDATION DES HYPOTHÈSES D'APPLICATION

Normalité de la distribution :
Etude graphique

Histogrammes pour la variable Year



VALIDATION DES HYPOTHÈSES D'APPLICATION

Normalité de la distribution :Test de Shapiro-Wilk

```
> shapiro.test(data_developed$Life.expectancy)

        Shapiro-Wilk normality test

data:  data_developed$Life.expectancy
W = 0.98486, p-value = 0.1662

> shapiro.test(data_developing$Life.expectancy)

        Shapiro-Wilk normality test

data:  data_developing$Life.expectancy
W = 0.94412, p-value = 4.72e-05
```

```
> shapiro.test(data_2000$Life.expectancy)

        Shapiro-Wilk normality test

data:  data_2000$Life.expectancy
W = 0.83148, p-value = 4.693e-07

> shapiro.test(data_2005$Life.expectancy)

        Shapiro-Wilk normality test

data:  data_2005$Life.expectancy
W = 0.89463, p-value = 5.049e-05

> shapiro.test(data_2010$Life.expectancy)

        Shapiro-Wilk normality test

data:  data_2010$Life.expectancy
W = 0.95535, p-value = 0.02116

> shapiro.test(data_2015$Life.expectancy)

        Shapiro-Wilk normality test

data:  data_2015$Life.expectancy
W = 0.93757, p-value = 0.002928
```

VALIDATION DES HYPOTHÈSES D'APPLICATION

- **Egalité des variances**

```
data$Year <- factor(data$Year, levels = c("2000","2005","2010","2015"))
```

```
leveneTest(Life.expectancy~Status,data=data)
leveneTest(Life.expectancy~Year,data=data)
```

```
> leveneTest(Life.expectancy~Year,data=data)
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   3  0.3031 0.8231
      252
```

```
> leveneTest(Life.expectancy~Status,data=data)
Levene's Test for Homogeneity of Variance (center = median)
       Df F value    Pr(>F)
group   1  48.849 2.427e-11 ***
      254
```

ANOVA

- Différence entre l'espérance de vie des pays développés et ceux en voie de développement
- Evolution de l'espérance de vie dans le monde de l'année 2000 à l'année 2015.

```
> kruskal.test(Life.expectancy~Year,data = data)

        Kruskal-Wallis rank sum test

data:  Life.expectancy by Year
Kruskal-Wallis chi-squared = 10.087, df = 3, p-value = 0.01784
```

```
> t.test(data_developed$Life.expectancy,data_developing$Life.expectancy)

        Welch Two Sample t-test

data:  data_developed$Life.expectancy and data_developing$Life.expectancy
t = 14.833, df = 171.52, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 11.03635 14.42459
sample estimates:
mean of x mean of y
 79.06250  66.33203
```

MERCI POUR VOTRE ATTENTION