Aide-mémoire pour l'utilisation de RStudio — Par Alexandre Michel

1 Quelques fonctions de base

Packages

```
install.packages("package")
library(package)
round(chiffre, 2)
liste = c(1,2,3,4,5)
```

2 Importation d'une base de données

Package WDI de la Banque Mondiale

```
base = WDI(country = c("US","CA"), indicator = c("SP.POP.TOTL"),
           start = 1960, end = 2020)
```

Fichier CSV

```
base = read.csv2(file.choose(), header = TRUE, encoding = "UTF-8")
base = read.csv(file.choose(), header = TRUE, encoding = "UTF-8")
```

Fichier Excel

```
base = read.xlsx(file.choose(), sheetIndex = 1,
                 header = TRUE, encoding = "UTF-8")
```

Fichier SPSS

```
base = spss.get(file.choose(), use.value.labels = TRUE, encoding = "UTF-8")
```

3 Manipulations générales sur une base

Infos sur la base

```
str(base)
```

Infos sur les variables

```
summary(base)
```

Renommer des variables

```
base = rename(base, replace = c("ancien_nom1" = "nouveau_nom1",
                                "ancien_nom2" = "nouveau_nom2", etc.))
```

Recoder les valeurs d'une variable

```
base$variable = recode(base$variable, "c(1) = 0 ; c(2,3,4,6) = 1 ; else = NA")
base$variable[base$variable == 99] = NA
```

Créer une nouvelle variable

```
base$agecat[base$age >= 18 & base$age <= 39] = 1
base$agecat[base$age >= 40 & base$age <= 59] = 2
base$agecat[base$age >= 60] = 3
```

Filtrer une base de données

```
base2 = subset(base, base$variable == "CA")
base2 = subset(base, select = c("variable1", "variable2", "variable3", etc.))
```

4 Analyse univariée

Variable qualitative

```
freq(base$variable)
```

Variable quantitative

```
moyenne = mean(base$variable, na.rm = TRUE)
mediane = median(base$variable, na.rm = TRUE)
ecart_type = sd(base$variable, na.rm = TRUE)
min = min(base$variable, na.rm = TRUE)
max = max(base$variable, na.rm = TRUE)
```

5 Analyse bivariée

Qualitative + Qualitative

```
t1 = table(base$variableY, base$variableX)
CrossTable(base$variableY, base$variableY,
           prop.t = FALSE, prop.chisq = FALSE)
chisq.test(t1)
assocstats(t1)
```

Qualitative + Quantitative

```
t2 = tapply(base$variableY, base$variableX, FUN = mean, na.rm = TRUE)
x = barplot(t2, xlab = "Groupes", ylab = "Moyennes",
            main = "Titre du graphique", ylim = c(0,max(t2+t2/10)))
text(x, t2+(t2/20), labels = as.character(round(t2, 2)))
```

Si la variable qualitative est dichotomique

```
t.test(base$variableY ~ base$variableX)
cohensD(base$variableY ~ base$variableX)
```

Si la variable qualitative est polytomique

```
cor.test(base$variableX, base$variableY, method = "kendall")
```

Quantitative + Quantitative

```
plot(base$variableX, base$variableY)
modele = lm(base$variableY ~ base$variableX)
summary(modele)
abline(modele)
cor.test(base$variableX, base$variableY, method = "pearson")
```