

PACS2021__1

UOC

Introducció

Les PACs es basaran en una base de dades obtinguda a partir del repositori de microdades del “Banc Mundial” a <https://microdata.worldbank.org/index.php/catalog/424/get-microdata>

Conté indicacions, entre d’altres de

1. *City* = Nom de la ciutat
2. *Country* = País
3. *Population2000* = Població de la ciutat a l’any 2000.
4. *PM10Concentration1999* = *PM10 concentrations (micro grams per cubic meter) in residential areas of cities larger than 100,000*, l’any 1999
5. *Region* = Classificació en regió geogràfica
6. *IncomeGroup* = Classificació segons nivell d’ingressos del país.

Per importar les dades podem usar la següent instrucció

```
dadesPM10<-read.table("AirPollution2000WB_UOC2.CSV", header=TRUE,
  sep=";",na.strings="NA",
  fileEncoding = "UTF-8", quote = "\"",
  colClasses=c(rep("character",4),rep("numeric",2),
    rep("character",2)))
```

PAC1

Un cop importades les dades,

1. a) Trobeu els resums numèrics de la variable **Population2000**. (10%)
b) Trobeu els resums numèrics de la variable **Population2000** a les ciutats d’Algeria i feu l’histograma corresponent. Comenteu els resultats. (20%)

a)

```
summary(dadesPM10$Population2000)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      6972  157530  264030   614693  547988 20951320
```

```
sd (dadesPM10$Population2000)
```

```
## [1] 1249855
```

b)

```
summary(dadesPM10$Population2000[dadesPM10$Country=="Algeria"])
```

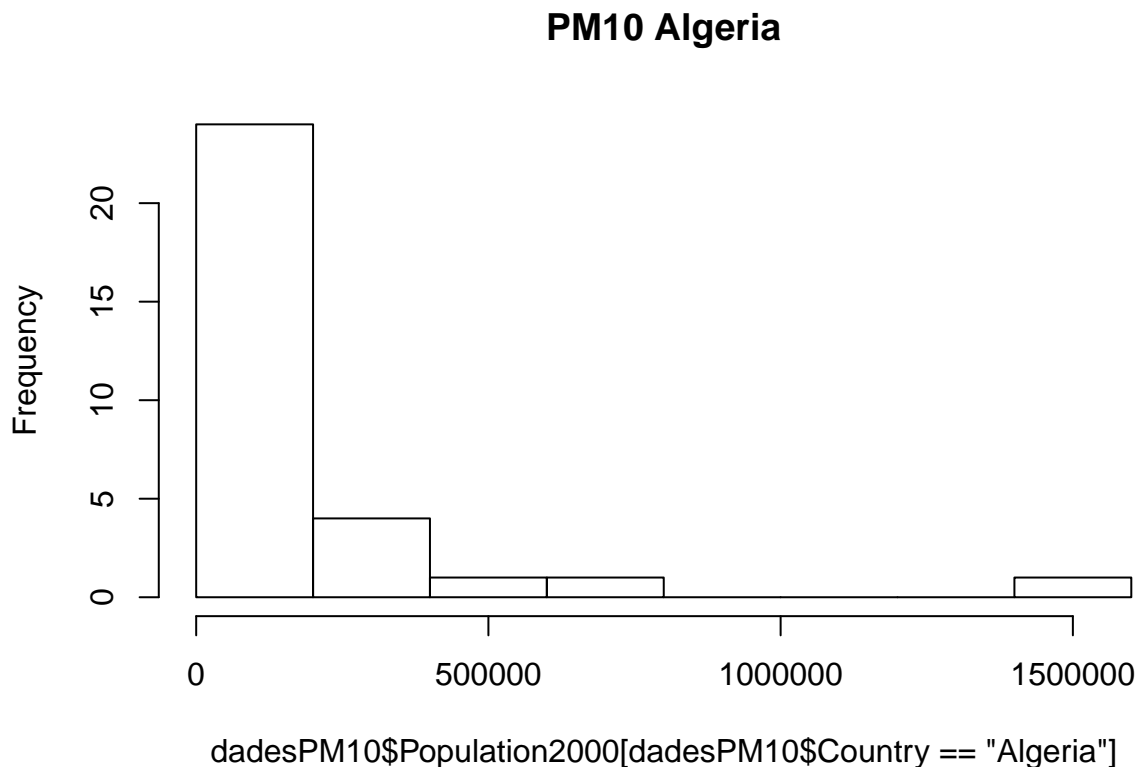
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  108075  116324  150614  229504  186557 1574799
```

```
sd(dadesPM10$Population2000[dadesPM10$Country=="Algeria"])
```

```
## [1] 279947.1
```

Hi ha molta asymetria amb una llarga cua cap a la dreta; la mitjana està desplaçada respecte a la mediana i es unimodal. La majoria de les ciutats estudiades tenen menys de 200000 habitants el mínim es de 108100 i el màxim de 1575000.

```
hist(dadesPM10$Population2000[dadesPM10$Country=="Algeria"],main="PM10 Algeria")
```



2. Trobeu el valor mínim de la variable Population2000 i el país i la ciutat corresponent. (15%)

```
min(dadesPM10$Population2000)
```

```
## [1] 6972
```

```
dadesPM10[which.min(dadesPM10$Population2000),  
           c("City", "Country", "Population2000")]
```

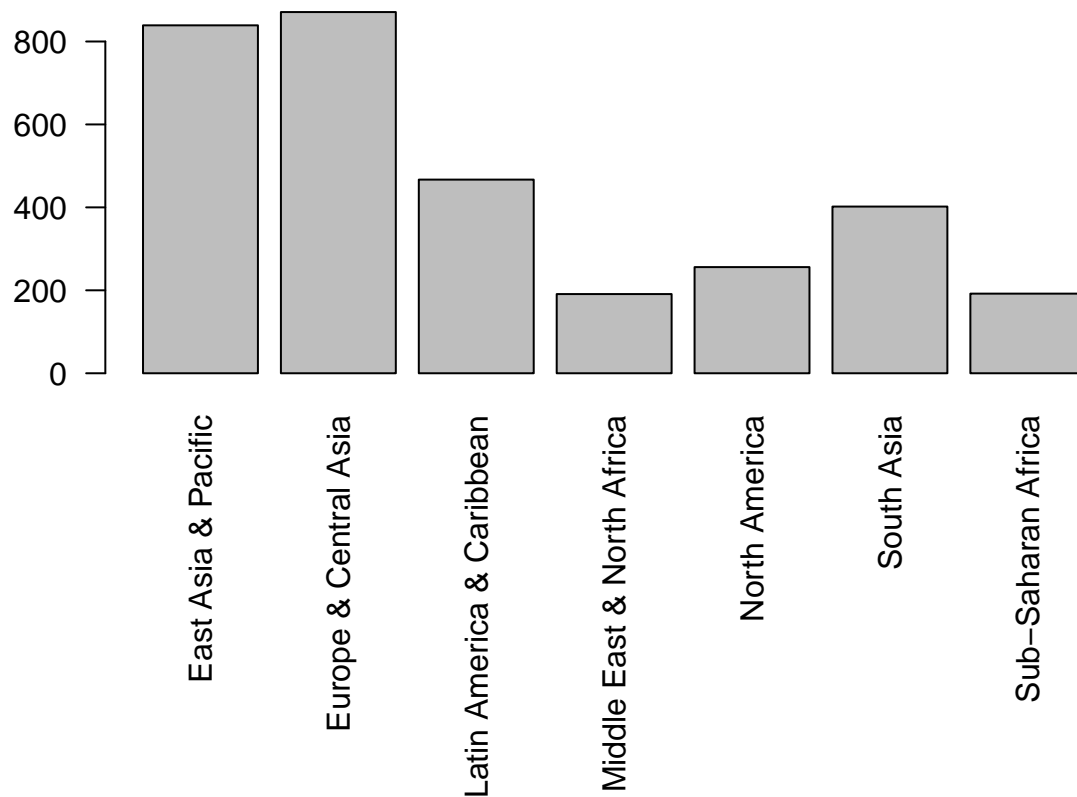
```
##      City      Country Population2000  
## 1908 VADUZ Liechtenstein          6972
```

3. Feu una taula i el diagrama de barres de la variable Region amb les etiquetes verticals i ampliant l'àrea del gràfic perquè hi càpiguen. Comenteu els resultats. (30%)

```
regions<-table(dadesPM10$Region)  
regions
```

```
##  
##      East Asia & Pacific      Europe & Central Asia  
##              839              871  
## Latin America & Caribbean Middle East & North Africa  
##              467              191  
##      North America      South Asia  
##              256              402  
##      Sub-Saharan Africa  
##              192
```

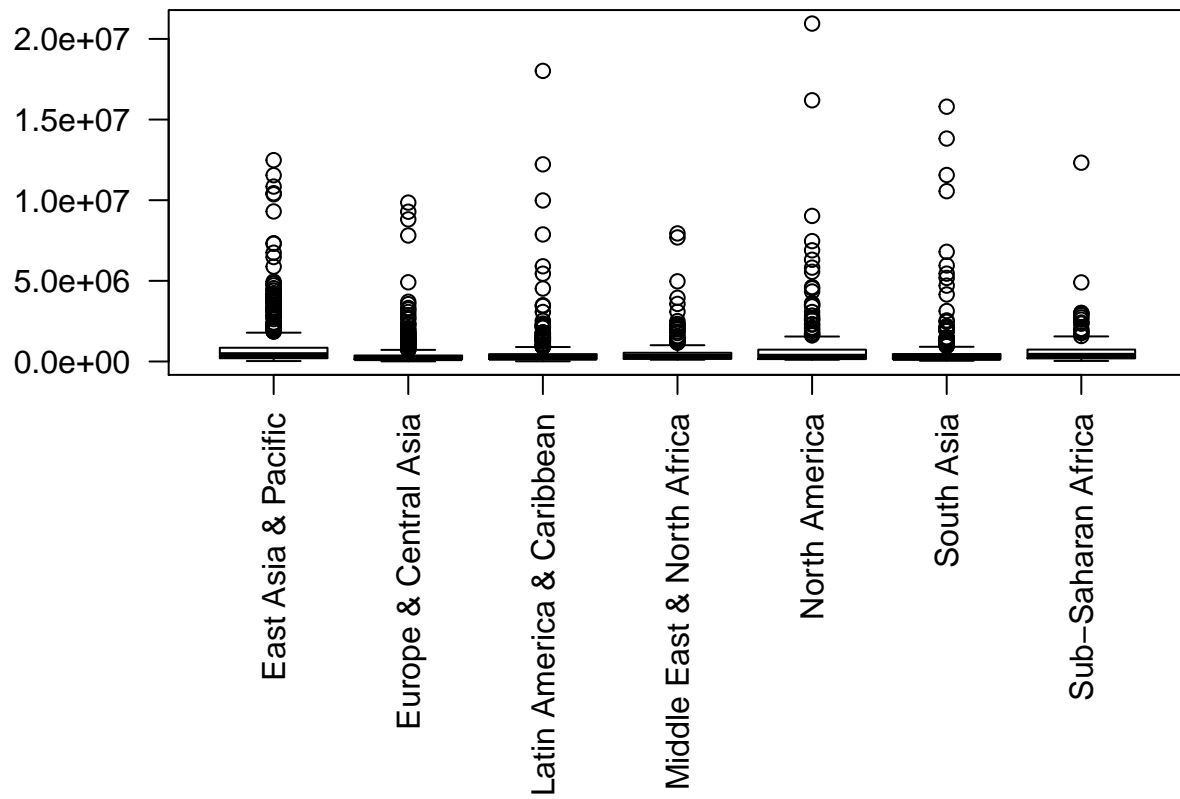
```
par(mar=c(12,5,1,1))  
barplot(regions,las=2)
```



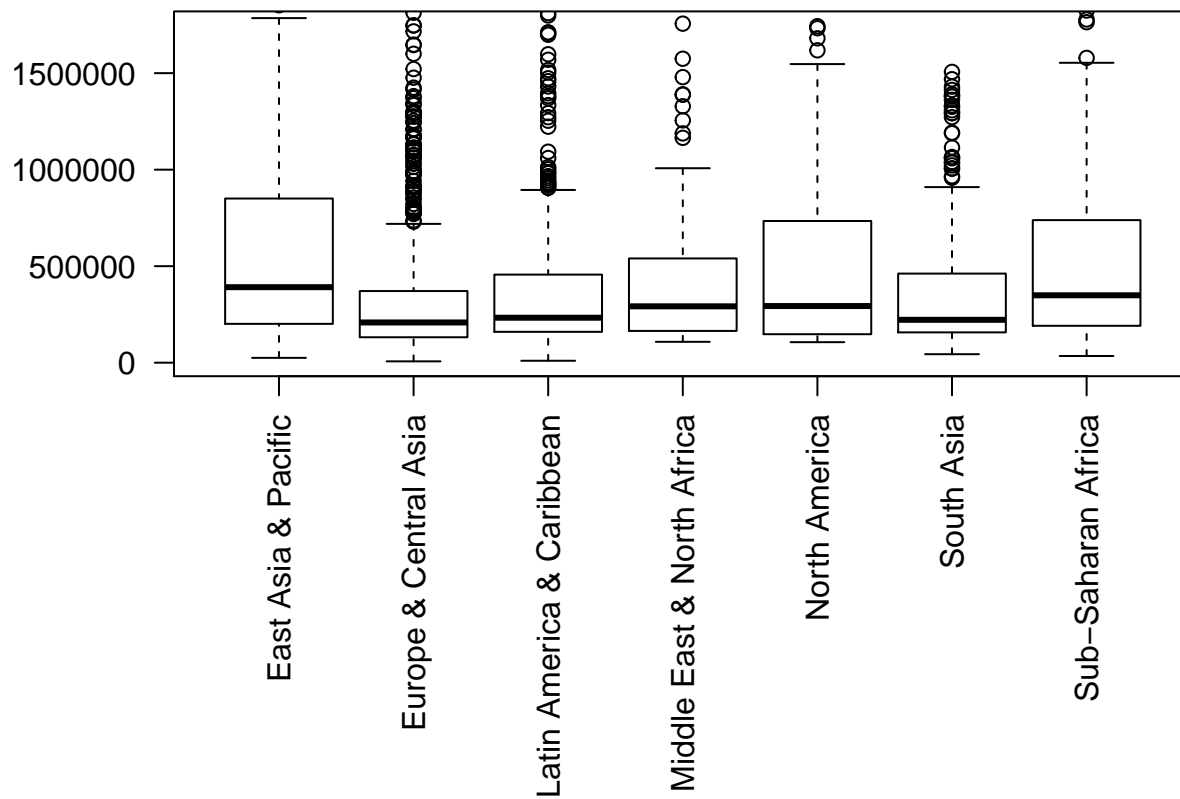
Hi ha més ciutats en l'estudi de les zones de l'Est-Pacífic i Europa-Àsia Central. De la zona d'Àfrica s'han agafat molt poques ciutats en aquest estudi.

4. Feu un Boxplot de Population2000 distingint per regions geogràfiques, amb les etiquetes verticals (podeu usar l'opció `las=2` al boxplot i ampliant l'àrea del gràfic perquè es veguin les caixes bé (instruccions `par=mar=c(12,5,1,1)` i `ylim=c(0.0e+00,1.75e+06)`). Comenteu els resultats.(25%)

```
par(mar=c(12,5,1,1))
boxplot(dadesPM10$Population2000~dadesPM10$Region,las=2,ylab="",xlab="")
```



```
boxplot(dadesPM10$Population2000~dadesPM10$Region,ylim=c(0.0e+00,1.75e+06),las=2,
        ylab="",xlab="")
```



Europa i Àsia Central tenen valors més baixos i menys dispersos. En l'altre extrem està el Pacífic i l'Est d'Àsia amb valors més alts i més dispersos.