

Tipologia i cicle de vida de les dades · Pràctica 2

Neteja i Anàlisi de Dades: iPhone, iPad and Macbook Index

Alejandro Tortosa Molla i Sergi Poy Garcia

1. Descripció del dataset

Utilitzem els datasets que vam extreure a la pràctica anterior, del Iphone Ipad and Macbook Index. Per a poder realitzar l'Índex haurem de fusionar-los e integrar les dades. Igual que el Big Mac Index volem tenir una comparativa de paritat de poder adquisitiu pero utilitzant com a referència els products d'Apple, per a poder tenir com a referència també un producte icònic, que no sigui del mercat de la restauració sinó de la tecnologia, recordem que Apple representa una part molt important de la venda de mòbils a nivell global amb el seu mòbil insígnia iPhone (<https://gs.statcounter.com/vendor-market-share/mobile>).

2. Integració i selecció

Per al nostre projecte el que farem serà treballar amb els 3 datasets que vam obtenir en la pràctica anterior:

- `average_monthly_salary_index.csv`: Aquí tenim el salari mitjà mensual en euros o en dolars i classificat per país.
- `world_currencies.csv`: Tenim l'abreviatura de totes les monedes i la seva equivalència en euros o dolars.
- `preus_mac_ipad_iphone.csv`: Finalment en aquest arxiu tenim totes les dades de tots els models de Macbook, iPad i iPhone amb el seu identificatiu. Tenim a més el país on es ven cada model, així com una columna amb al divisa en la que está el preu de cada dispositiu.

Obtenim així totes les dades importants i representatives a l'hora de voler elaborar l'índex com a objectiu de la pràctica. Però per a aconseguir l'objectiu les haurem de fusionar en un sol dataset. L'arxiu troncal a partir del qual treballarem és el dels `preus_mac_ipad_iphone.csv`:

Primer fusionem `preus_mac_ipad_iphone.csv` amb `world_currencies.csv` a partir de la columna `Currency`. Després fusionem el dataset resultant amb `average_monthly_salary_index.csv` a partir de la columna `Country`.

Això però ho farem després de la neteja.

3. Neteja de les dades

La neteja de les dades va ítegrament relacionada amb la integració d'aquestes, ja que per a poder-les fusionar amb èxit em de veure que els valors entre columnes coincideixin. Per a més tard quan aquesta integració es faci amb èxit puguem reduir la quantitat de columnes i veure si tenim una cohesió entre elles.

Per tant anem a treballar amb les dades i les carreguem a R:

```
preus_csv <- read.csv("csv/preus_mac_ipad_iphone.csv")
salaris_csv <- read.csv("csv/average_monthly_salary_index.csv")
divises_csv <- read.csv("csv/world_currencies.csv")
```

La columna de "price" té aquest format: "{fullPrice: 499990.0}". Hem d'extreure amb `gsub` i una `RegEx` el preu. Com que la columna original es de tipus caràcter (`chr`) modifiquem amb `as.numeric` el seu format per obtenir el preu en numèric.

```
class(preus_csv$price)
```

```
## [1] "character"
```

```
preus_csv$price <- as.numeric(gsub("[[:digit:]]", "", preus_csv$price))
head(preus_csv)
```

```
##      sku price category      name currency country
## 1 MGN63 499990      mac MacBook Air with M1 chip - Space Gray      HUF      hu
## 2 MGND3 499990      mac      MacBook Air with M1 chip - Gold      HUF      hu
## 3 MGN93 499990      mac      MacBook Air with M1 chip - Silver      HUF      hu
## 4 MLY23 749990      mac MacBook Air with M2 chip - Starlight      HUF      hu
## 5 MLY13 599990      mac MacBook Air with M2 chip - Starlight      HUF      hu
## 6 MLY03 749990      mac      MacBook Air with M2 chip - Silver      HUF      hu
##              store
## 1 AOS: Hungary Consumer
## 2 AOS: Hungary Consumer
## 3 AOS: Hungary Consumer
## 4 AOS: Hungary Consumer
## 5 AOS: Hungary Consumer
## 6 AOS: Hungary Consumer
```

També hem d'extreure el país de la columna “store” amb una RegEx:

```
# En aquest cas gsub serà suficient per substituir el valor per el coincident amb la RegEx.
preus_csv$store <- gsub("^AOS: (.*) [Consumer|consumer].*$", "\\1", preus_csv$store)
unique(preus_csv$store)
```

```
## [1] "Hungary"      "Philippines"
## [3] "German"       "Canada"
## [5] "Japan"        "Austrian"
## [7] "Hong Kong Chinese" "Brazil"
## [9] "Spain"        "Czech"
## [11] "Korea"        "French Canada"
## [13] "Finnish"      "Irish"
## [15] "Turkey"      "NZ"
## [17] "US"          "Belgium French"
## [19] "Poland"      "Thailand"
## [21] "Mexico"      "Denmark"
## [23] "Swedish"     "Singapore"
## [25] "UAE"         "Belgium Flemish"
## [27] "IN"          "Australia"
## [29] "UK"          "Swiss French"
## [31] "Malaysia"    "Norway"
## [33] "Italian"     "Taiwan"
## [35] "AOS: Apple Store Luxembourg" "Swiss German"
## [37] "China"       "French"
## [39] "HongKong"    "Portugal"
## [41] "Dutch"
```

Com podem veure pels valors únics el filtratge no ha sortit del tot bé, i no coincideixen amb els equivalents de `salaries_csv` per a fer una fusió exitosa. Per exemple de valor tenim “UK” però a `salaries_csv` tenim “United Kingdom”, per tant utilitzarem la funció `case_when()` per a modificar-los.

Aparentment es un treball tediós però així ens asegurem que les dades estiguin ven integrades i tinguin coherència entre elles. Farem una petita trampa, afegir “dup” a un dels camps dels països amb més d’un idioma, per a desfer-nos de les dades repetides (trobem països amb dues pàgines web per tenir dues idiomes, aquestes les haurem de treure per ser dades duplicades)

```
preus_csv$Country <- case_when(
  preus_csv$store == "German" ~ "Germany",
  preus_csv$store == "Austrian" ~ "Austria",
  preus_csv$store == "Hong Kong Chinese" ~ "dup",
  preus_csv$store == "Czech" ~ "Czech Republic",
  preus_csv$store == "Korea" ~ "South Korea",
  preus_csv$store == "French Canada" ~ "dup",
  preus_csv$store == "Finnish" ~ "Finland",
  preus_csv$store == "Irish" ~ "Ireland",
  preus_csv$store == "NZ" ~ "New Zealand",
  preus_csv$store == "US" ~ "United States",
  preus_csv$store == "Belgium French" ~ "Belgium",
  preus_csv$store == "Swedish" ~ "Sweden",
  preus_csv$store == "UAE" ~ "United Arab Emirates",
  preus_csv$store == "Belgium Flemish" ~ "dup",
  preus_csv$store == "IN" ~ "India",
  preus_csv$store == "UK" ~ "United Kingdom",
  preus_csv$store == "Swiss French" ~ "Switzerland",
  preus_csv$store == "Italian" ~ "Italy",
  preus_csv$store == "AOS: Apple Store Luxembourg" ~ "Luxembourg",
  preus_csv$store == "Swiss German" ~ "dup",
  preus_csv$store == "French" ~ "France",
  preus_csv$store == "HongKong" ~ "Hong Kong",
  preus_csv$store == "Dutch" ~ "Netherlands",
  TRUE ~ preus_csv$store
  # Amb TRUE ~ preus_csv$store copiem els valors que no hem especificat a la nova columna.
)
head(preus_csv)
```

```
##      sku price category      name currency country
## 1 MGN63 499990      mac MacBook Air with M1 chip - Space Gray    HUF      hu
## 2 MGN63 499990      mac MacBook Air with M1 chip - Gold      HUF      hu
## 3 MGN93 499990      mac MacBook Air with M1 chip - Silver    HUF      hu
## 4 MLY23 749990      mac MacBook Air with M2 chip - Starlight    HUF      hu
## 5 MLY13 599990      mac MacBook Air with M2 chip - Starlight    HUF      hu
## 6 MLY03 749990      mac MacBook Air with M2 chip - Silver    HUF      hu
##      store Country
## 1 Hungary Hungary
## 2 Hungary Hungary
## 3 Hungary Hungary
## 4 Hungary Hungary
## 5 Hungary Hungary
## 6 Hungary Hungary
```

Donada la neteja anterior fem una fusió dels 3 datasets:

```
# Eliminem les files duplicades (països amb pàgines amb dues idiomes, mateix preu)
preus_csv <- distinct(preus_csv)
```

```

preus_csv <- preus_csv[preus_csv$Country != "dup", ]

# Primer el dataset principal de preus amb el de canvi de divises. Posem all.x per a assegurar-nos que
# les pàgines de Apple
preus_divises_csv <- merge(preus_csv, divises_csv, by.x = "currency", by.y = "Currency", all.x = TRUE)

#Segon el dataset anterior amb el de salaris mensuals mitjans per país.
apple_index_raw <- merge(preus_divises_csv, salaris_csv, by.x = "Country", by.y = "Country", all.x = TRUE)

# Mirem com queda el dataset
head(apple_index_raw)

```

```

##      Country currency  sku price category
## 1 Australia      AUD MPQ83   999      ipad
## 2 Australia      AUD MNXP3  1899      ipad
## 3 Australia      AUD  Z12T  2499      mac
## 4 Australia      AUD MM6V3  1249      ipad
## 5 Australia      AUD MNXX3  3129      ipad
## 6 Australia      AUD MPUR3  1399      iphone
##
##                                name country      store
## 1                10.9-inch iPad Wi-Fi 256GB - Silver      au Australia
## 2                12.9-inch iPad Pro Wi-Fi 128GB - Space Gray      au Australia
## 3                24-inch Yellow iMac with 4.5K Retina display      au Australia
## 4 10.9-inch iPad Air Wi-Fi + Cellular 64GB - Starlight      au Australia
## 5                12.9-inch iPad Pro Wi-Fi 1TB - Silver      au Australia
## 6                iPhone 14 128GB Starlight      au Australia
##  X1.EUR.in.Currency X1.USD.in.Currency EUR.Average.Monthly.Salary
## 1                1.55                1.50                3915.54
## 2                1.55                1.50                3915.54
## 3                1.55                1.50                3915.54
## 4                1.55                1.50                3915.54
## 5                1.55                1.50                3915.54
## 6                1.55                1.50                3915.54
##  USD.Average.Monthly.Salary
## 1                4050.82
## 2                4050.82
## 3                4050.82
## 4                4050.82
## 5                4050.82
## 6                4050.82

```

```

# Mirem si existeixen valors nuls
anyNA(apple_index_raw)

```

```
## [1] TRUE
```

Observem com existeixen valors nuls, així que passarem a veure com tractar-los

```

# Com que ens apareixen valors nuls, veurem quins son
summary(apple_index_raw)

```

```
##      Country      currency      sku      price
```

```
## Length:6697      Length:6697      Length:6697      Min.   :    449
## Class :character  Class :character  Class :character  1st Qu.:   1539
## Mode  :character  Mode  :character  Mode  :character  Median :   4049
##                                     Mean  :  88473
##                                     3rd Qu.:  25999
##                                     Max.   :4710000
##
##      category      name      country      store
## Length:6697      Length:6697      Length:6697      Length:6697
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
## X1.EUR.in.Currency X1.USD.in.Currency EUR.Average.Monthly.Salary
## Length:6697      Length:6697      Min.   : 280.6
## Class :character  Class :character  1st Qu.: 978.1
## Mode  :character  Mode  :character  Median :2294.5
##                                     Mean  :2147.0
##                                     3rd Qu.:3039.1
##                                     Max.   :5519.1
##                                     NA's   :181
## USD.Average.Monthly.Salary
## Min.   : 290.3
## 1st Qu.:1011.9
## Median :2373.7
## Mean   :2221.1
## 3rd Qu.:3144.1
## Max.   :5709.8
## NA's   :181
```

Observem col els valors nuls es troben a “EUR.Average.Monthly.Salary” i “USD.Average.Monthly.Salary”, així que veurem si trobem el motiu i com ho podem arreglar.

```
# Com que ens apareixen valors nuls, veurem qhins son
head(apple_index_raw[which(is.na(apple_index_raw$EUR.Average.Monthly.Salary)) , ])
```

```
##      Country currency  sku  price category
## 3078 Luxembourg    EUR MPQC3  762.92    ipad
## 3079 Luxembourg    EUR MJV83 1405.74     mac
## 3080 Luxembourg    EUR MGPH3 1618.47     mac
## 3081 Luxembourg    EUR MQ323 1913.58    iphone
## 3082 Luxembourg    EUR Z132 1618.47     mac
## 3083 Luxembourg    EUR Z133 1840.87     mac
##                                     name country
## 3078      10.9-inch iPad Wi-Fi 256GB - Pink    lu
## 3079  24-inch Green iMac with 4.5K Retina display    lu
## 3080  24-inch Green iMac with 4.5K Retina display    lu
## 3081      iPhone 14 Pro 1TB Deep Purple    lu
## 3082 24-inch Orange iMac with 4.5K Retina display    lu
## 3083 24-inch Orange iMac with 4.5K Retina display    lu
##                                     store X1.EUR.in.Currency X1.USD.in.Currency
```

```
## 3078 AOS: Apple Store Luxembourg      1.00      0.97
## 3079 AOS: Apple Store Luxembourg      1.00      0.97
## 3080 AOS: Apple Store Luxembourg      1.00      0.97
## 3081 AOS: Apple Store Luxembourg      1.00      0.97
## 3082 AOS: Apple Store Luxembourg      1.00      0.97
## 3083 AOS: Apple Store Luxembourg      1.00      0.97
##      EUR.Average.Monthly.Salary USD.Average.Monthly.Salary
## 3078      NA      NA
## 3079      NA      NA
## 3080      NA      NA
## 3081      NA      NA
## 3082      NA      NA
## 3083      NA      NA
```

Com podem veure es deu a que de l'arxiu de `salaries_csv` no s'han descarregat dades per a luxembourg. Per a tractar aquest NA el que farem serà veure si podem trobar la dada, i en eixe cas introduir-la manualment. Si mirem a la mateixa pàgina de la que ens hem descarregat les dades, i busquem per país (<https://www.numbeo.com/cost-of-living/in/Luxembourg>), veurem com la dada existeix, així que la introduïrem:

```
# Introduïm la dada manualment
apple_index_raw$'EUR.Average.Monthly.Salary'[apple_index_raw$Country == "Luxembourg"] <- 4140.31

summary(apple_index_raw)
```

```
##      Country      currency      sku      price
## Length:6697      Length:6697      Length:6697      Min.   :   449
## Class :character      Class :character      Class :character      1st Qu.:  1539
## Mode  :character      Mode  :character      Mode  :character      Median :   4049
##                                     Mean   :  88473
##                                     3rd Qu.: 25999
##                                     Max.   :4710000
##
##      category      name      country      store
## Length:6697      Length:6697      Length:6697      Length:6697
## Class :character      Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character      Mode  :character
##
##
##
##      X1.EUR.in.Currency X1.USD.in.Currency EUR.Average.Monthly.Salary
## Length:6697      Length:6697      Min.   : 280.6
## Class :character      Class :character      1st Qu.: 978.4
## Mode  :character      Mode  :character      Median :2296.1
##                                     Mean   :2200.8
##                                     3rd Qu.:3100.6
##                                     Max.   :5519.1
##
##      USD.Average.Monthly.Salary
## Min.   : 290.3
## 1st Qu.:1011.9
## Median :2373.7
## Mean   :2221.1
```

```
## 3rd Qu.:3144.1
## Max.    :5709.8
## NA's    :181
```

Veiem que ara ja no tenim NA a les dades en EUR, que son les que utilitzarem per a crear l'index.

Veiem però les columnes X1.EUR.in.Currency i X1.USD.in.Currency estan en "chr", per tant les hem de transformar, a més alguns valors tenen "," i no punt, per tant farem una substitució amb gsub:

```
apple_index_raw$X1.EUR.in.Currency <- as.numeric(gsub(",", "", apple_index_raw$X1.EUR.in.Currency))
apple_index_raw$X1.USD.in.Currency <- as.numeric(gsub(",", "", apple_index_raw$X1.USD.in.Currency))

apple_index_raw$X1.EUR.in.Currency <- as.numeric(apple_index_raw$X1.EUR.in.Currency)
apple_index_raw$X1.USD.in.Currency <- as.numeric(apple_index_raw$X1.USD.in.Currency)

head(apple_index_raw)
```

```
##      Country currency  sku price category
## 1 Australia      AUD MPQ83   999      ipad
## 2 Australia      AUD MNXP3  1899      ipad
## 3 Australia      AUD Z12T  2499      mac
## 4 Australia      AUD MM6V3  1249      ipad
## 5 Australia      AUD MNXX3  3129      ipad
## 6 Australia      AUD MPUR3  1399      iphone
##
##              name country      store
## 1          10.9-inch iPad Wi-Fi 256GB - Silver      au Australia
## 2          12.9-inch iPad Pro Wi-Fi 128GB - Space Gray      au Australia
## 3          24-inch Yellow iMac with 4.5K Retina display      au Australia
## 4 10.9-inch iPad Air Wi-Fi + Cellular 64GB - Starlight      au Australia
## 5          12.9-inch iPad Pro Wi-Fi 1TB - Silver      au Australia
## 6          iPhone 14 128GB Starlight      au Australia
##      X1.EUR.in.Currency X1.USD.in.Currency EUR.Average.Monthly.Salary
## 1                1.55                1.5      3915.54
## 2                1.55                1.5      3915.54
## 3                1.55                1.5      3915.54
## 4                1.55                1.5      3915.54
## 5                1.55                1.5      3915.54
## 6                1.55                1.5      3915.54
##      USD.Average.Monthly.Salary
## 1                4050.82
## 2                4050.82
## 3                4050.82
## 4                4050.82
## 5                4050.82
## 6                4050.82
```

També tenim moltes columnes que ja no ens fan falta, com la de sku, country (la segona) i store. La resta encara poden ser rellevant, per tant eliminem les mencionades:

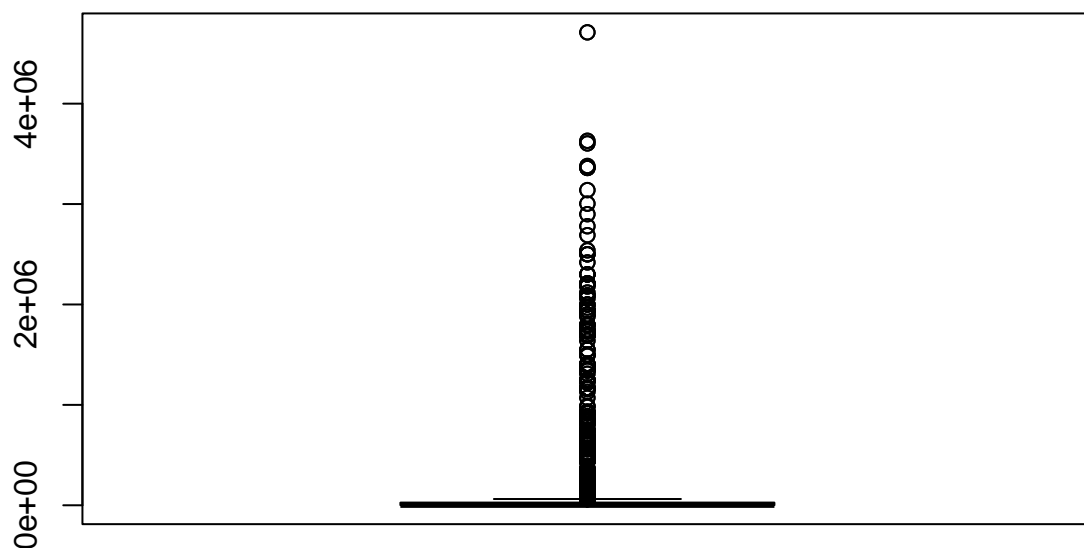
```
apple_index_raw <- apple_index_raw[, -c(3, 7, 8)]
head(apple_index_raw)
```

```
##      Country currency price category
```

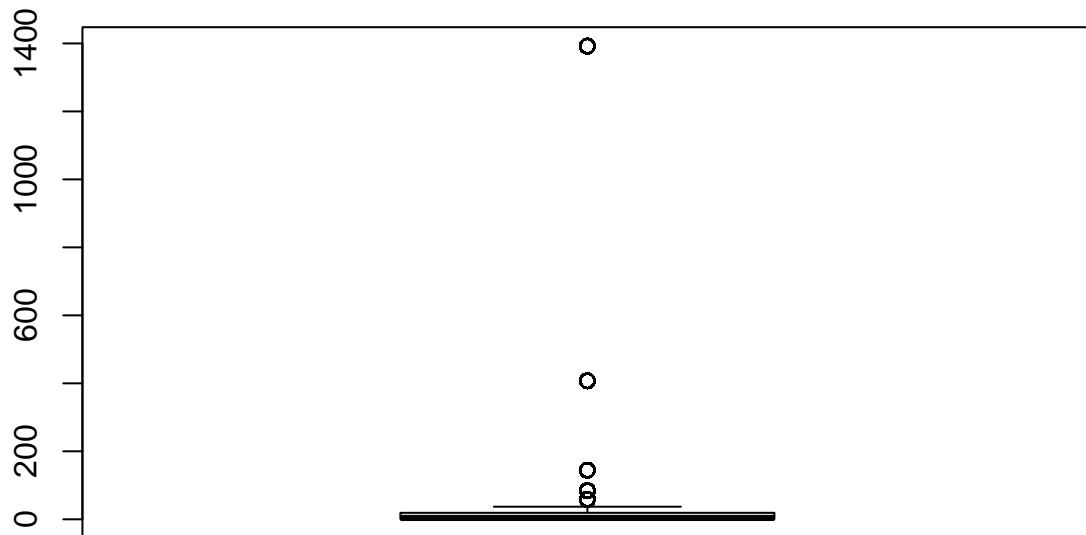
```
## 1 Australia      AUD    999    ipad
## 2 Australia      AUD   1899    ipad
## 3 Australia      AUD   2499    mac
## 4 Australia      AUD   1249    ipad
## 5 Australia      AUD   3129    ipad
## 6 Australia      AUD   1399    iphone
##
##                                name X1.EUR.in.Currency
## 1                10.9-inch iPad Wi-Fi 256GB - Silver      1.55
## 2                12.9-inch iPad Pro Wi-Fi 128GB - Space Gray 1.55
## 3                24-inch Yellow iMac with 4.5K Retina display 1.55
## 4 10.9-inch iPad Air Wi-Fi + Cellular 64GB - Starlight      1.55
## 5                12.9-inch iPad Pro Wi-Fi 1TB - Silver      1.55
## 6                iPhone 14 128GB Starlight                  1.55
##  X1.USD.in.Currency EUR.Average.Monthly.Salary USD.Average.Monthly.Salary
## 1                1.5                3915.54                4050.82
## 2                1.5                3915.54                4050.82
## 3                1.5                3915.54                4050.82
## 4                1.5                3915.54                4050.82
## 5                1.5                3915.54                4050.82
## 6                1.5                3915.54                4050.82
```

Finalment mirarem si tenim valors extrems al dataset i finalitzarem així el procés de neteja i integració. Principalment evaluarem els valors extrems per veure si la fusió s'ha realitzat amb èxit:

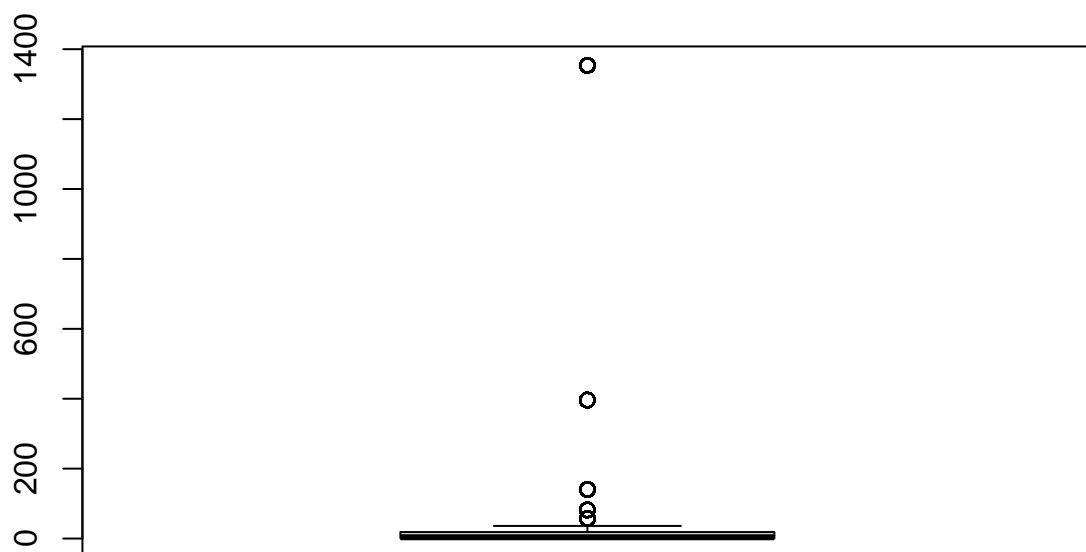
```
# Evaluem els boxplots de les columnes fusionades, així veurem si s'ha fusionat bé:
boxplot(apple_index_raw$price)
```



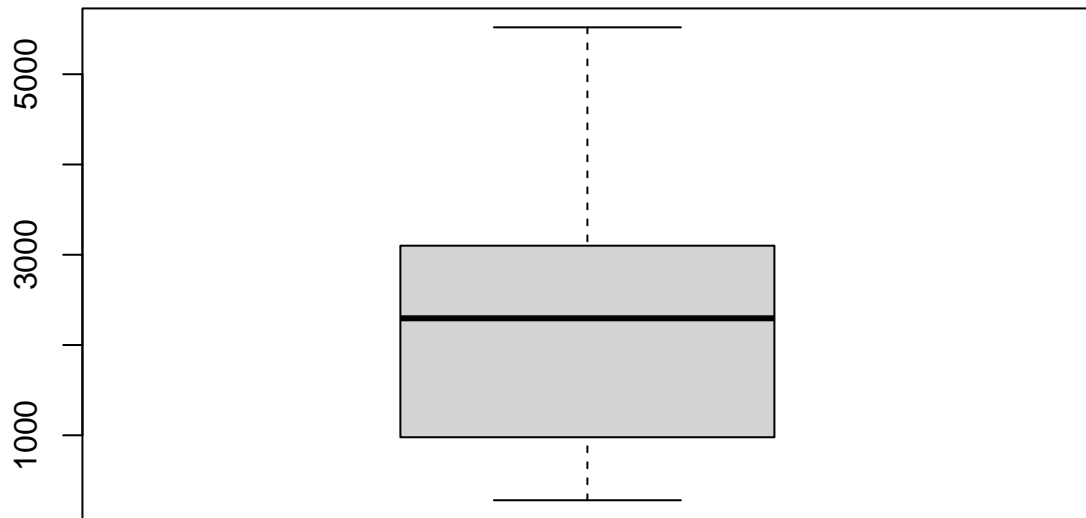

```
boxplot(apple_index_raw$X1.EUR.in.Currency)
```



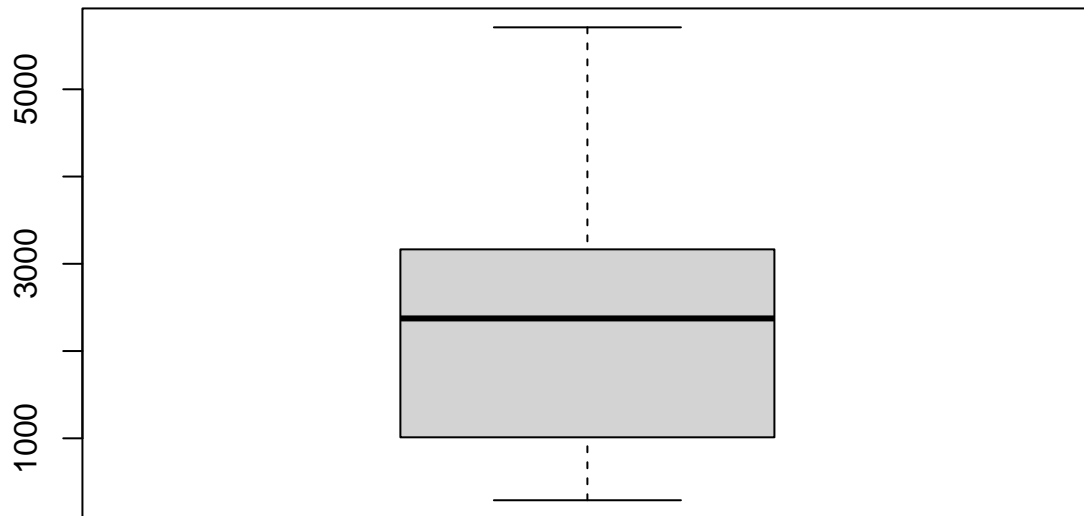
```
boxplot(apple_index_raw$X1.USD.in.Currency)
```



```
boxplot(apple_index_raw$EUR.Average.Monthly.Salary)
```



```
boxplot(apple_index_raw$USD.Average.Monthly.Salary)
```



```
apple_index_raw[which.max(apple_index_raw$price),]
```

```
##      Country currency  price category      name
## 4903 South Korea      KRW 4710000      mac 16-inch MacBook Pro - Silver
##      X1.EUR.in.Currency X1.USD.in.Currency EUR.Average.Monthly.Salary
## 4903          1392.09          1353.83          2105.13
##      USD.Average.Monthly.Salary
## 4903          2177.86
```

```
apple_index_raw[which.max(apple_index_raw$X1.EUR.in.Currency),]
```

```
##      Country currency  price category      name
## 4888 South Korea      KRW 1700000      iphone iPhone 14 512GB (PRODUCT)RED
##      X1.EUR.in.Currency X1.USD.in.Currency EUR.Average.Monthly.Salary
## 4888          1392.09          1353.83          2105.13
##      USD.Average.Monthly.Salary
## 4888          2177.86
```

```
apple_index_raw[which.max(apple_index_raw$X1.USD.in.Currency),]
```

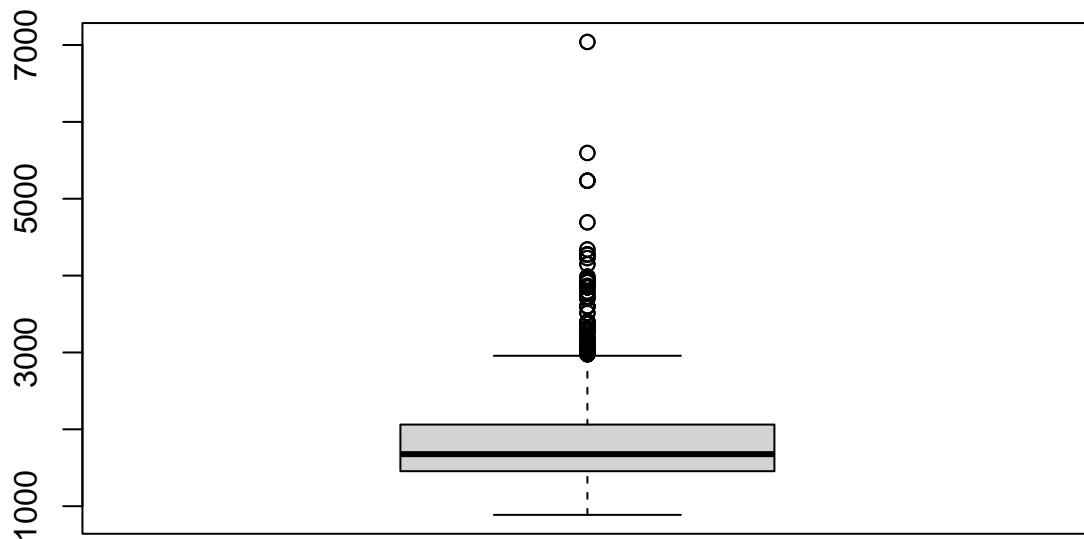
```
##      Country currency  price category      name
## 4888 South Korea      KRW 1700000      iphone iPhone 14 512GB (PRODUCT)RED
##      X1.EUR.in.Currency X1.USD.in.Currency EUR.Average.Monthly.Salary
## 4888          1392.09          1353.83          2105.13
```

```
##      USD.Average.Monthly.Salary
## 4888                2177.86
```

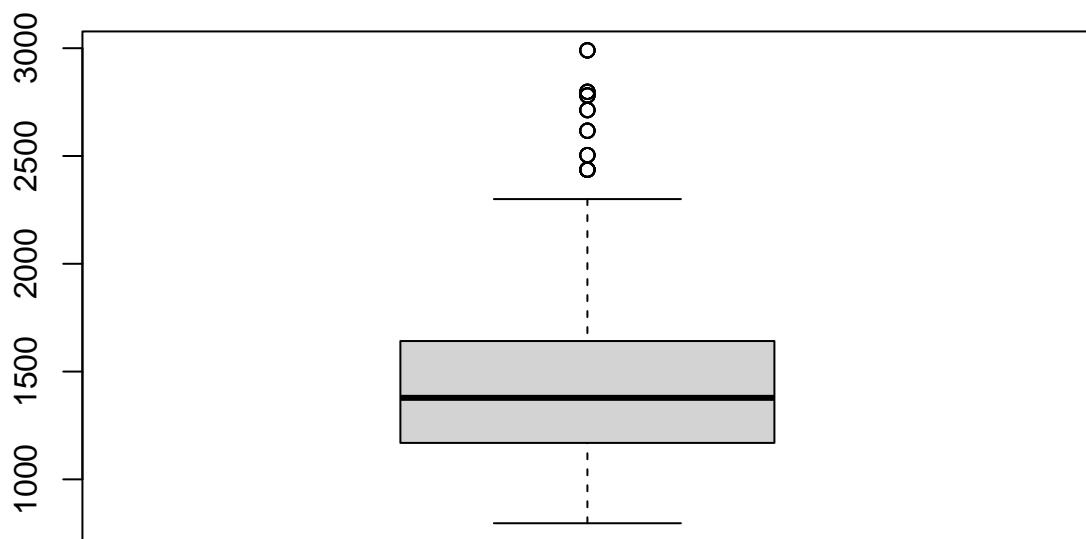
Cal dir que en el nostre cas és senzill si coneixem el dataset, com podem veure els valors extrems es corresponen als preus en won coreans, una divisa on 1€ correspon a 1,392.09₩.

També veurem si existeixen valors extrems als preus dels productes convertits a EUR

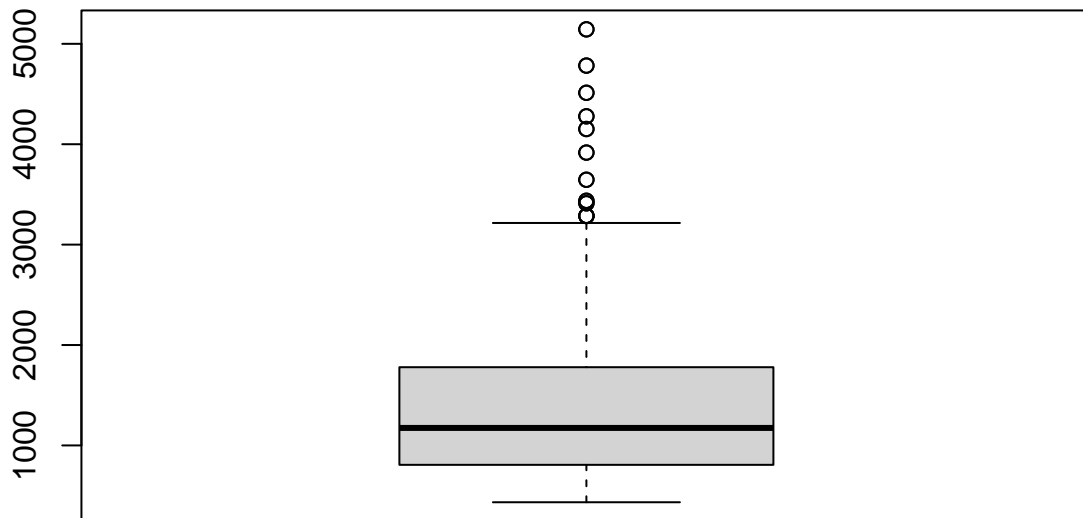
```
# Evaluem els boxplots de les columnes fusionades, així veurem si s'ha fusionat bé:
apple_index_raw$price_eur <- apple_index_raw$price / apple_index_raw$X1.EUR.in.Currency
boxplot(apple_index_raw[apple_index_raw$category == "mac", c(10)])
```



```
boxplot(apple_index_raw[apple_index_raw$category == "iphone", c(10)])
```



```
boxplot(apple_index_raw[apple_index_raw$category == "ipad", c(10)])
```



Observem com a la primera gràfica que és la dels “Mac”, el nombre de preus en EUR que semblen valors extrems és superior. Després, els dels “Ipad” i “Iphone” son semblants. En tots els casos, com són dades descarregades de la pàgina web de la companyia en cada país, assumirem que son correctes i no les modificarem.

4. Anàlisi de les dades.

5. Representació dels resultats.

L'objectiu del nostre projecte es crear un índex per països on es compari quants mesos de mitja ha de treballar una persona per comprar els diferents productes d'Apple, per tant haurem de fer un índex per almenys un producte de cada, per l'iPhone, l'iPad i el MacBook.

Primer hem de calcular els mesos que em de treballar per comprar-ne un per país: Amb l'equivalència de l'Euro creem la fórmula que ens dona els mesos que hem de treballar per a poder comprar un producte Apple en determinat país. La fórmula és: $\text{Preu producte apple} / (\text{Sou mitjà mensual} * \text{Canvi de divisa a 1€})$

```
apple_index_raw$months_to_buy_product <- apple_index_raw$price / (apple_index_raw$EUR.Average.Monthly.S
head(apple_index_raw)
```

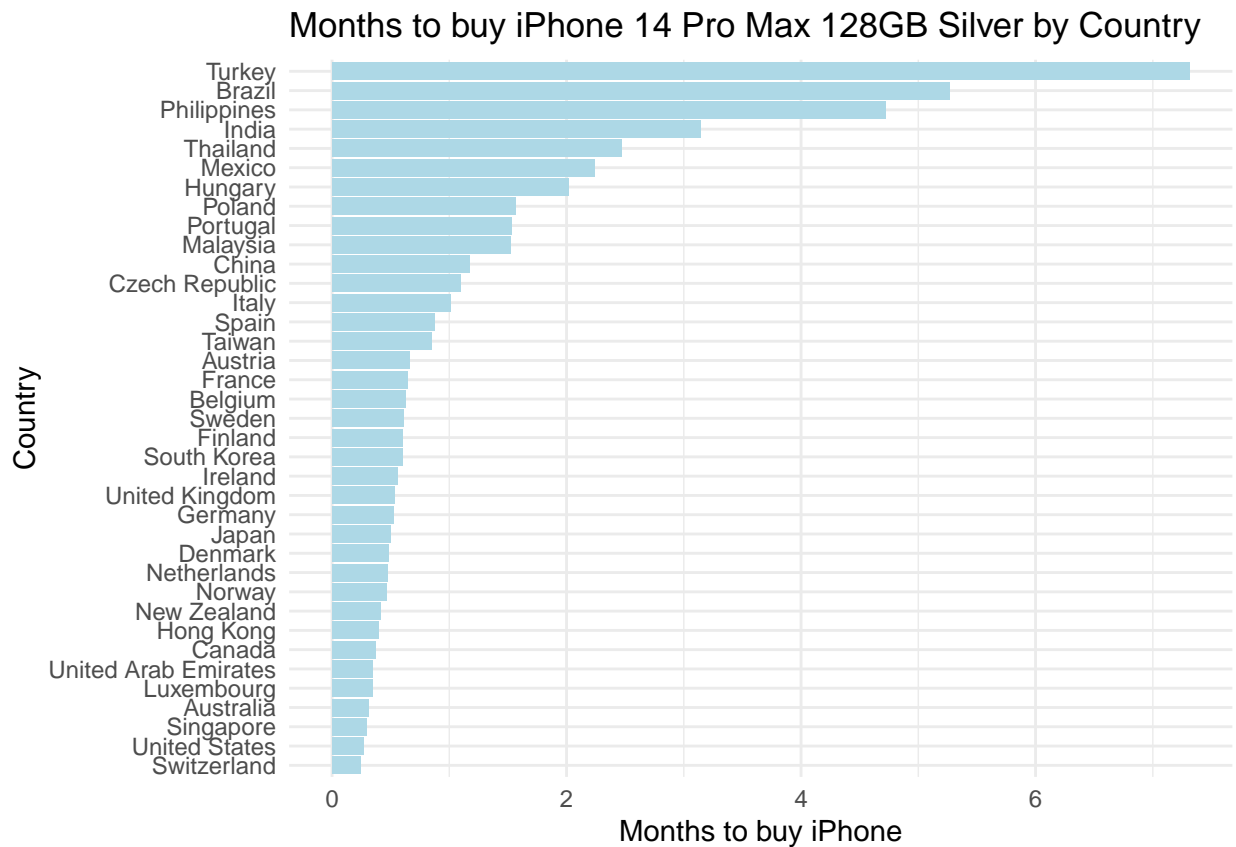
```
##      Country currency price category
## 1 Australia      AUD   999      ipad
## 2 Australia      AUD  1899      ipad
## 3 Australia      AUD  2499       mac
## 4 Australia      AUD  1249      ipad
```

```
## 5 Australia      AUD  3129    ipad
## 6 Australia      AUD  1399    iphone
##                                     name X1.EUR.in.Currency
## 1                10.9-inch iPad Wi-Fi 256GB - Silver      1.55
## 2                12.9-inch iPad Pro Wi-Fi 128GB - Space Gray 1.55
## 3                24-inch Yellow iMac with 4.5K Retina display 1.55
## 4 10.9-inch iPad Air Wi-Fi + Cellular 64GB - Starlight      1.55
## 5                12.9-inch iPad Pro Wi-Fi 1TB - Silver      1.55
## 6                iPhone 14 128GB Starlight                 1.55
##  X1.USD.in.Currency EUR.Average.Monthly.Salary USD.Average.Monthly.Salary
## 1                1.5                3915.54                4050.82
## 2                1.5                3915.54                4050.82
## 3                1.5                3915.54                4050.82
## 4                1.5                3915.54                4050.82
## 5                1.5                3915.54                4050.82
## 6                1.5                3915.54                4050.82
##  price_eur months_to_buy_product
## 1  644.5161          0.1646047
## 2 1225.1613          0.3128971
## 3 1612.2581          0.4117588
## 4  805.8065          0.2057970
## 5 2018.7097          0.5155635
## 6  902.5806          0.2305124
```

Finalment agafem un subset amb només els productes que coincideixen amb el model d'iPhone que volem analitzar i ho plotejem, fent l'índex de l'iPhone:

```
iphone14PM_index <- subset(apple_index_raw, apple_index_raw$name == "iPhone 14 Pro Max 128GB Silver")

ggplot(iphone14PM_index, aes(x=reorder(Country,months_to_buy_product), y=months_to_buy_product)) +
  geom_bar(stat="identity", fill = "lightblue") +
  xlab("Country") +
  ylab("Months to buy iPhone") +
  ggtitle("Months to buy iPhone 14 Pro Max 128GB Silver by Country") +
  coord_flip() +
  theme_minimal()
```

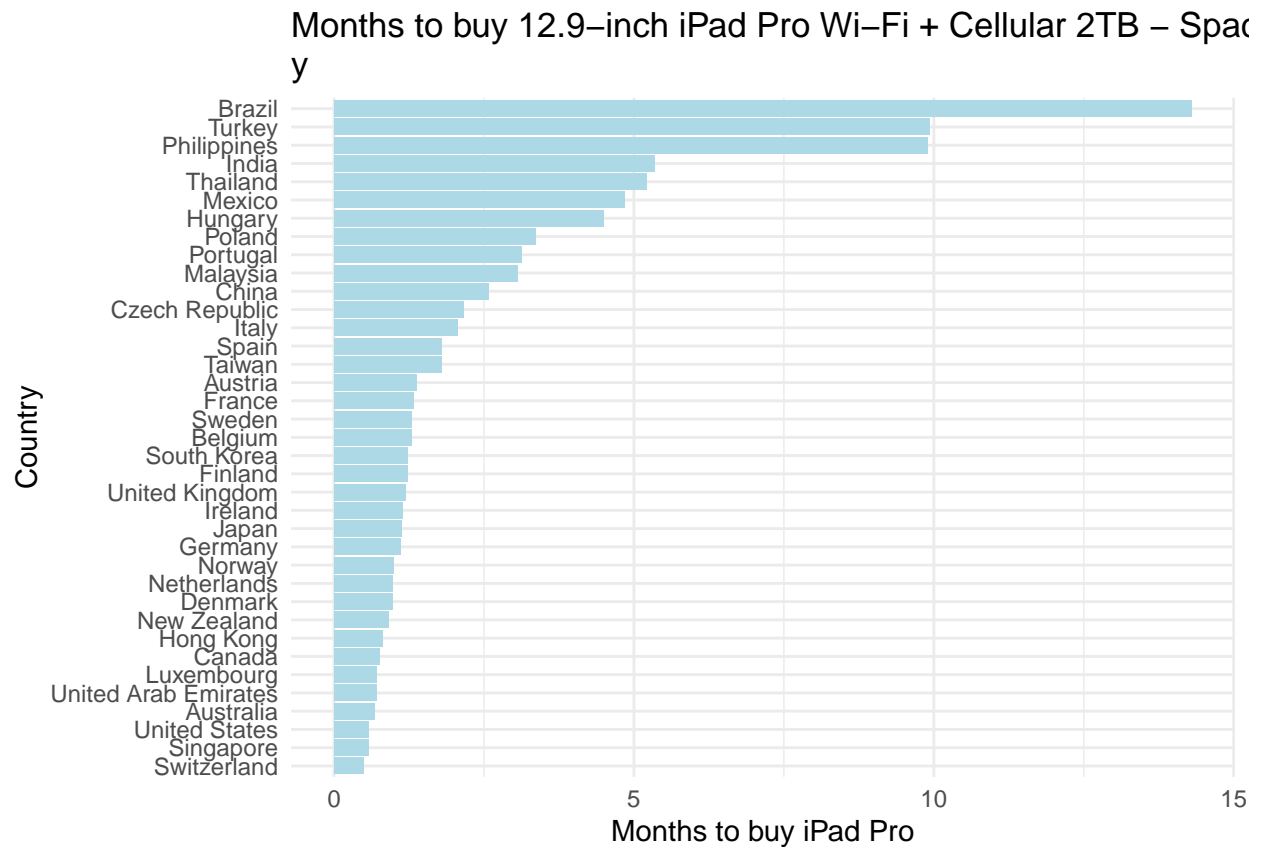
Ara fem el mateix per un model d'Ipad:

```

ipadPro12_index <- subset(apple_index_raw, apple_index_raw$name == "12.9-inch iPad Pro Wi-Fi + Cellular")

ggplot(ipadPro12_index, aes(x=reorder(Country,months_to_buy_product), y=months_to_buy_product)) +
  geom_bar(stat="identity", fill = "lightblue") +
  xlab("Country") +
  ylab("Months to buy iPad Pro") +
  ggtitle("Months to buy 12.9-inch iPad Pro Wi-Fi + Cellular 2TB - Space Gray") +
  coord_flip() +
  theme_minimal()

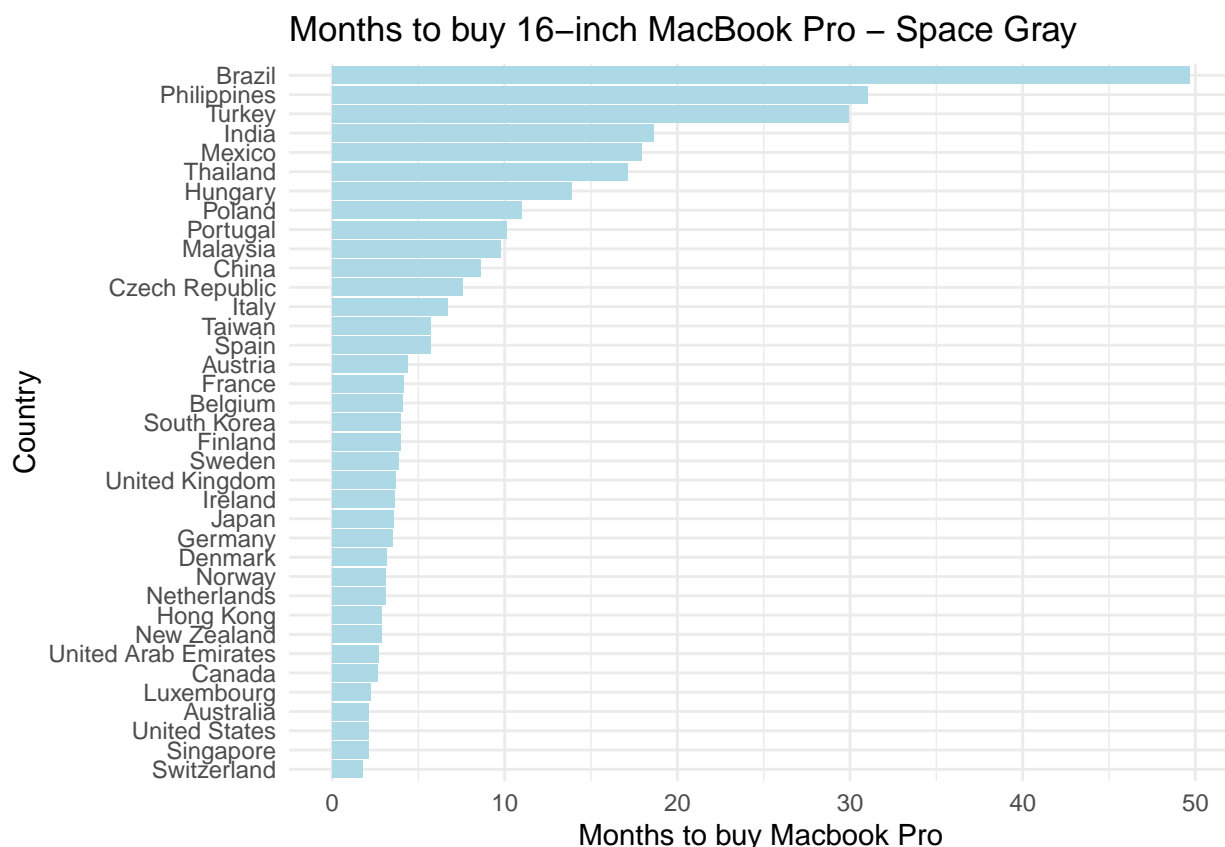
```



Finalment per a un model de Macbook:

```
MacBookPro16_index <- subset(apple_index_raw, apple_index_raw$name == "16-inch MacBook Pro - Space Gray")

ggplot(MacBookPro16_index, aes(x=reorder(Country,months_to_buy_product), y=months_to_buy_product)) +
  geom_bar(stat="identity", fill = "lightblue") +
  xlab("Country") +
  ylab("Months to buy Macbook Pro") +
  ggtitle("Months to buy 16-inch MacBook Pro - Space Gray") +
  coord_flip() +
  theme_minimal()
```



Observem com en general són els mateixos països els que necessiten mes mesos de treball en mitja per a comprar tant iPad, com iPhone i Mac, encara que en cada producte ballen les posicions. Passa el mateix amb els països als que menys esforç els suposa

6. Resolució del problema.

Com podem veure hem pogut crear aquests índex a partir de les dades d'un moment donat dels preus dels productes Apple, la paritat monetària del moment i el salari mensual mitjà per país. Per tant l'objectiu de crear un Índex igual que el Big Mac Índex ha estat assolit tot i que podríem haver estat molt més rigorosos amb la neteja de dades i podríem haver treballat moltes més gràfiques que podrien haver sortit de les dades. Tot i així trobem que la base del problema l'hem resolt.

7. Codi.

Com que hem treballat amb Rmarkdown hem pogut veure com estavem resolent el problema mentre ensenyaem quin codi hem utilitzat per a assolir-ho.

```
kable(data.frame(Contribucions = c("Investigació prèvia", "Redacció de les respostes", "Desenvolupament
```

Contribucions	Signatura
Investigació prèvia	Alex, Sergi
Redacció de les respostes	Alex, Sergi
Desenvolupament del codi	Alex, Sergi

Contribucions	Signatura
Participació al video	Alex, Sergi