Pràctica

Group 02

2022-03-18

```
if( !require(tidyverse)) install.packages('tidyverse')
```

```
## Loading required package: tidyverse
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.
3.1 ──
```

```
## ✓ ggplot2 3.3.5     ✓ purrr   0.3.4
## ✓ tibble  3.1.6     ✓ dplyr   1.0.8
## ✓ tidyr   1.2.0     ✓ stringr 1.4.0
## ✓ readr   2.1.2     ✓ forcats 0.5.1
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts()
 ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
if( !require(ggplot2)) install.packages('ggplot2')
if( !require(dplyr)) install.packages('dplyr')


library(tidyverse)
library(ggplot2)
library(dplyr)
library(forcats)
library(ggthemr)
ggthemr('fresh')

path <- file.path(getwd(), 'any_drinking.csv')
any_drinking=read.csv(path)
view(any_drinking)
```

```
any_drinking_ex1 <- any_drinking %>%
                    select(state, location, ends_with('2010'))
head(any_drinking_ex1, 5)
```

```
##      state       location both_sexes_2010 females_2010 males_2010
## 1 National  United States           56.1         49.3       63.2
## 2  Alabama        Alabama           42.5         35.3       50.1
## 3  Alabama Autauga County           42.5         34.4       50.9
## 4  Alabama Baldwin County           54.6         47.3       62.1
## 5  Alabama Barbour County           38.6         30.2       47.4
```

```
any_drinking_ex2 <- any_drinking %>%
                    select(state, location, ends_with('2010')) %>%
                    filter(males_2010 != 'NA', females_2010 != 'NA') %>%
                    mutate(DIF_SEXE_2010 =  abs(males_2010 - females_2010))


head(any_drinking_ex2, 10)
```

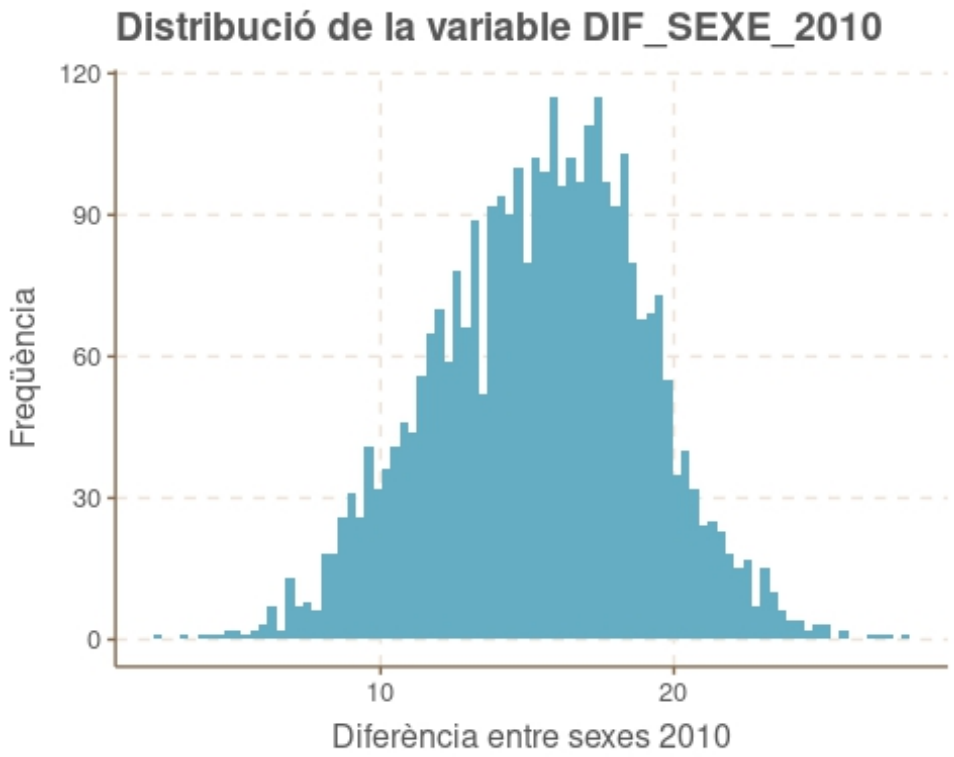
```
##      state       location both_sexes_2010 females_2010 males_2010
## 1  National  United States           56.1         49.3       63.2
## 2   Alabama        Alabama           42.5         35.3       50.1
## 3   Alabama Autauga County           42.5         34.4       50.9
## 4   Alabama Baldwin County           54.6         47.3       62.1
## 5   Alabama Barbour County           38.6         30.2       47.4
## 6   Alabama    Bibb County           34.4         26.5       42.6
## 7   Alabama  Blount County           33.5         24.7       42.6
## 8   Alabama Bullock County           37.6         28.0       47.7
## 9   Alabama  Butler County           35.2         26.4       44.4
## 10  Alabama Calhoun County           37.3         29.7       45.2
##   DIF_SEXE_2010
## 1          13.9
## 2          14.8
## 3          16.5
## 4          14.8
```

```
## 5         17.2
## 6         16.1
## 7         17.9
## 8         19.7
## 9         18.0
## 10        15.5
```

```
str(any_drinking_ex2)
```

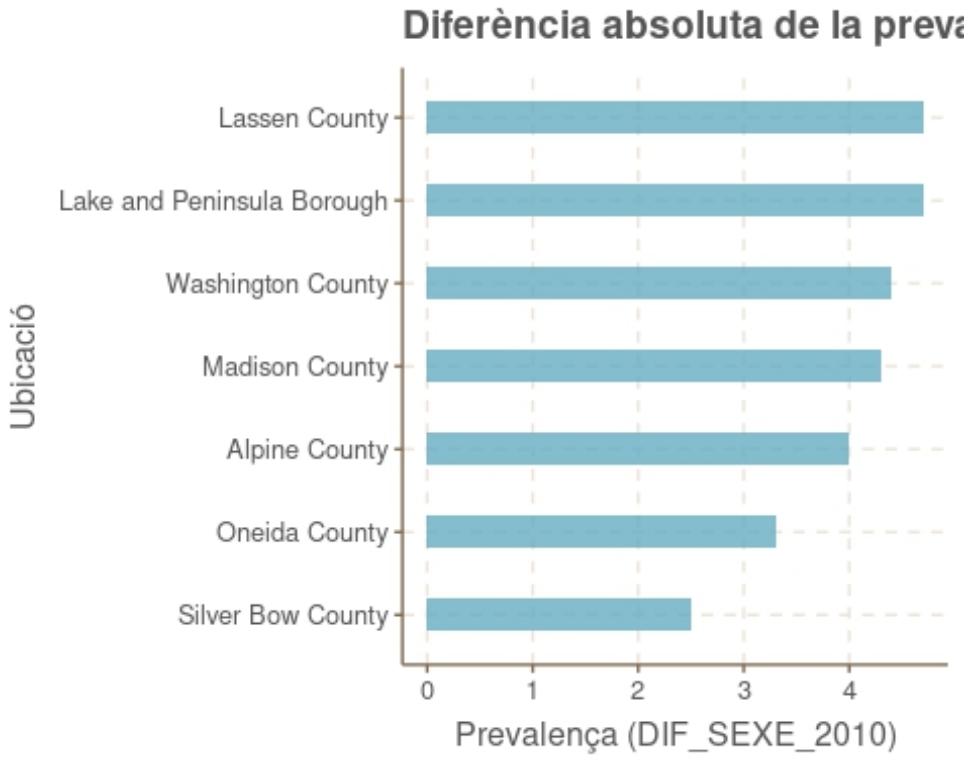
```
## 'data.frame':    3169 obs. of  6 variables:
##      state         : Factor w/ 62 levels 'Alabama','Alaska',..: 33 1 1 1 1 1 1 1 1 1 ...
##      location      : Factor w/ 1907 levels '','Abbeville County',..: 1747 11 87 94 105
 154 169 228 238 251 ...
##      both_sexes_2010: num  56.1 42.5 42.5 54.6 38.6 34.4 33.5 37.6 35.2 37.3 ...
##      females_2010  : num  49.3 35.3 34.4 47.3 30.2 26.5 24.7 28 26.4 29.7 ...
##      males_2010    : num  63.2 50.1 50.9 62.1 47.4 42.6 42.6 47.7 44.4 45.2 ...
##      DIF_SEXE_2010 : num  13.9 14.8 16.5 14.8 17.2 16.1 17.9 19.7 18 15.5 ...
```

```
ggthemr('fresh')
ggplot(any_drinking_ex2, aes(DIF_SEXE_2010)) + geom_histogram(binwidth = 0.3) + xl
ab('Diferència entre sexes 2010') + ylab('Freqüència') + ggtitle('Distribució de la variab
le DIF_SEXE_2010')
```

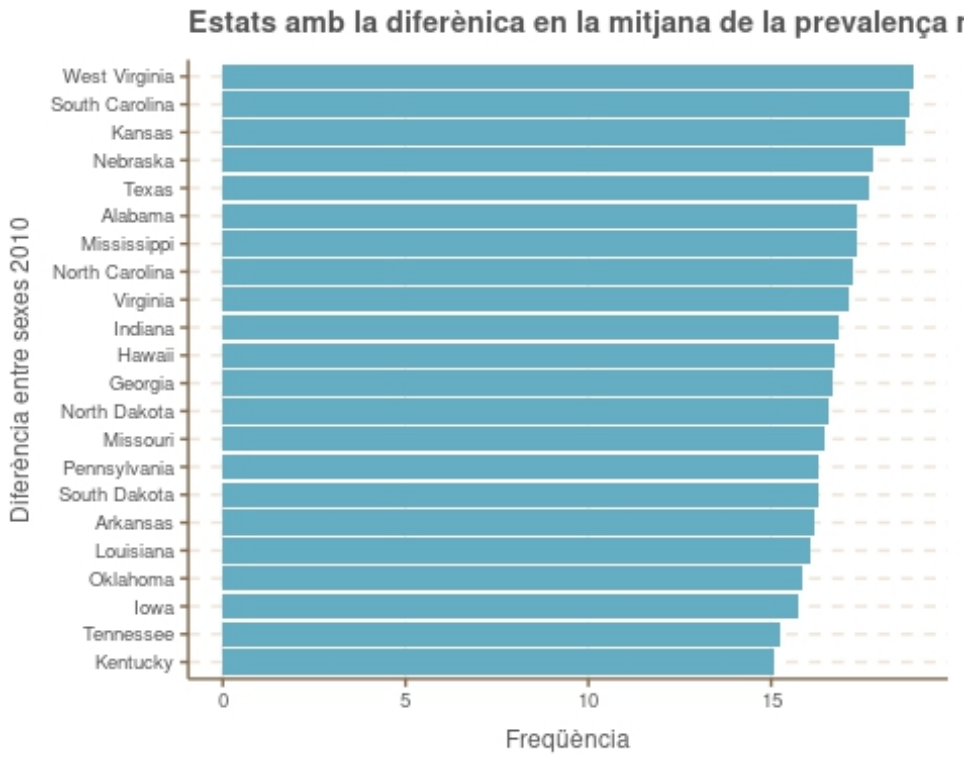



```
diff_less_than_5 = any_drinking_ex2 %>%
                   filter(DIF_SEXE_2010 <= 5)

diff_less_than_5 %>%
 mutate(location = fct_reorder(location, DIF_SEXE_2010)) %>%
                    ggplot(aes(DIF_SEXE_2010, location)) +
                      geom_bar(stat='identity', alpha=.8, width=.4) +
                      xlab('Prevalença (DIF_SEXE_2010)') + ggtitle('Diferència absoluta
de la prevalença') +
                      ylab('Ubicació')
```




```
any_drinking_ex2 %>%
                group_by(state) %>%
                summarise(mitjana = mean(DIF_SEXE_2010)) %>%
                filter(mitjana > 15) %>%
                mutate(state = fct_reorder(state, mitjana)) %>%
                  ggplot(aes(state, mitjana)) +
                  geom_col() + xlab('Diferència entre sexes 2010') + ylab('Freqüència') + c
oord_flip() + theme(text = element_text(size=9)) + ggtitle('Estats amb la diferènica en la
mitjana de la prevalença més gran a 15')
```



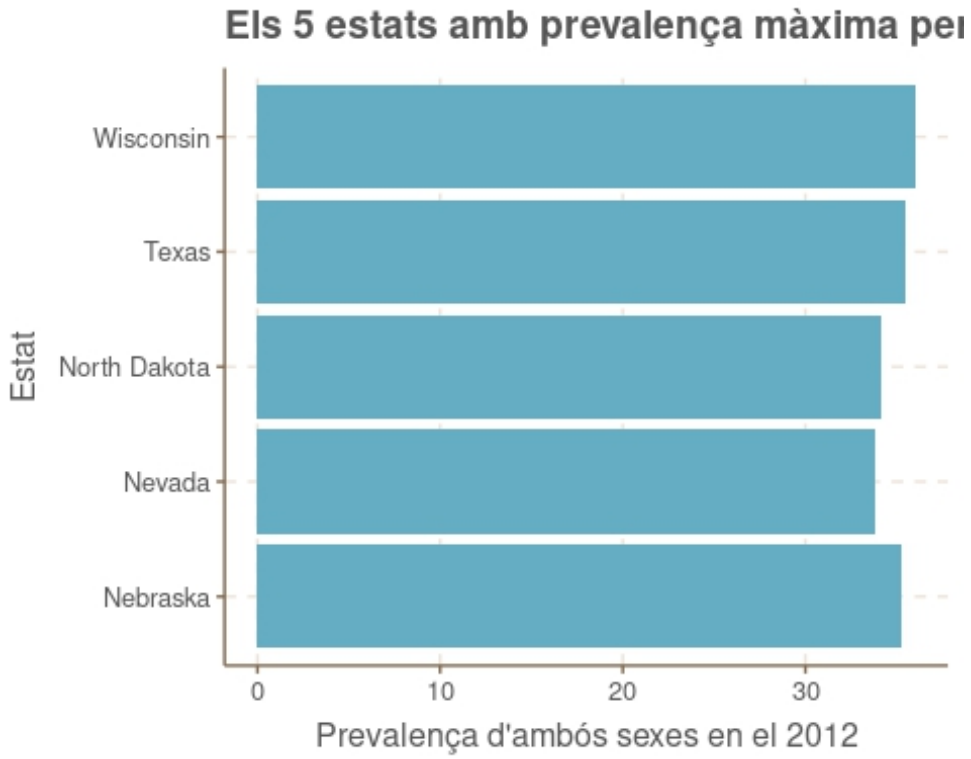

```
path <- file.path(getwd(), 'binge_drinking.csv')
binge_drinking = read.csv(path)
County_DATA = binge_drinking %>%filter(grepl('County',location))
head(County_DATA, 5)
```

```
##     state       location both_sexes_2002 females_2002 males_2002
## 1 Alabama Autauga County           13.6          6.6       20.8
## 2 Alabama Baldwin County           17.9          9.7       26.4
## 3 Alabama Barbour County           12.8          5.4       20.4
## 4 Alabama    Bibb County           11.3          4.4       18.5
## 5 Alabama  Blount County           10.2          4.1       16.5
##   both_sexes_2003 females_2003 males_2003 both_sexes_2004 females_2004
## 1            13.3          6.9       20.0            13.0          6.5
## 2            17.4          9.8       25.3            17.4          9.5
## 3            12.5          5.6       19.6            12.1          5.2
## 4            11.3          4.8       18.0            11.4          4.9
## 5            10.4          4.5       16.5            10.6          4.6
##   males_2004 both_sexes_2005 females_2005 males_2005 both_sexes_2006
```

```
## 1      19.8          12.8         6.8      18.9          12.1
## 2      25.7          16.8         9.6      24.3          16.6
## 3      19.4          11.9         5.4      18.6          10.7
## 4      18.1          11.0         5.1      17.2          10.5
## 5      16.9          10.5         4.8      16.4          10.0
##   females_2006 males_2006 both_sexes_2007 females_2007 males_2007
## 1          7.0       17.4           12.8          7.5       18.2
## 2         10.3       23.1           16.9         10.4       23.7
## 3          5.0       16.7           11.1          5.0       17.3
## 4          5.3       15.9           10.9          5.5       16.4
## 5          4.9       15.4           10.3          5.0       15.9
##   both_sexes_2008 females_2008 males_2008 both_sexes_2009 females_2009
## 1            13.1          7.3       19.2            13.0          7.4
## 2            17.7         10.5       25.2            17.0         10.0
## 3            12.0          5.2       19.1            11.5          5.0
## 4            11.6          5.7       17.8            11.7          5.6
## 5            10.8          4.8       16.9            10.5          4.6
##   males_2009 both_sexes_2010 females_2010 males_2010 both_sexes_2011
## 1       18.8          13.3          7.9       18.8          14.4
## 2       24.2          16.7          9.9       23.7          18.6
## 3       18.3          12.7          5.9       19.8          13.5
## 4       17.9          11.4          5.5       17.5          12.4
## 5       16.6          10.2          4.4       16.2          11.3
##   females_2011 males_2011 both_sexes_2012 females_2012 males_2012
## 1          8.8       20.1           13.2          7.9       18.7
## 2         11.7       25.8           16.9         10.4       23.7
## 3          6.3       21.0           12.4          5.4       19.6
## 4          6.3       18.8           11.4          5.7       17.4
## 5          5.2       17.6           10.3          4.6       16.2
```

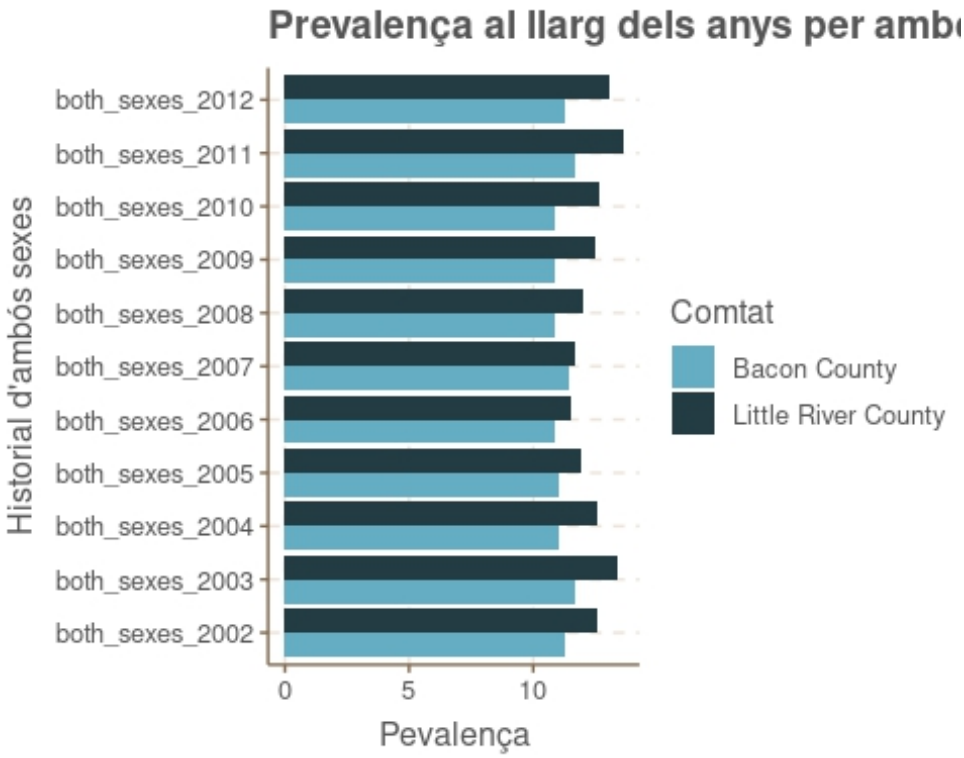
```
County_DATA %>%
            select(state, both_sexes_2012) %>%
            arrange(desc(both_sexes_2012)) %>%
```

```
        head(5) %>%
        ggplot(aes(state, both_sexes_2012)) + geom_col() + coord_flip() + xlab('Estat') + ylab('Prevalença d'ambós sexes en el 2012') + ggtitle('Els 5 estats amb prevalença màxima per ambós sexes en 2012')
```




```
df8 = County_DATA %>% select(location, starts_with('both_sexes'))
df8 = df8 %>% filter(location == 'Little River County' | location == 'Bacon County')
df9 = df8 %>% gather('both_sexes_historical', 'prevalença', 2:12)

ggplot(df9,aes(fill=location,both_sexes_historical,prevalença)) + geom_col(position='dodge') + coord_flip() + ylab('Pevalença') + xlab('Historial d'ambós sexes') + scale_fill_discrete('Comtat') + ggtitle('Prevalença al llarg dels anys per ambdós sexes')
```


Prevalença al llarg dels anys per ambd
Historial d'ambós sexes
both_sexes_2012
both_sexes_2011
both_sexes_2010
both_sexes_2009
both_sexes_2008
both_sexes_2007
both_sexes_2006
both_sexes_2005
both_sexes_2004
both_sexes_2003
both_sexes_2002
0
5
10
Pevalença
Comtat
Bacon County
Little River County