

Publikációkész grafika R-ben

Abari Kálmán

2024. április 22.

Tartalomjegyzék

Előszó	5
1 Bevezetés	6
2 Szükséges csomagok	8
3 Hisztogram és barátai	9
3.1 Eloszlás egy csoportra	9
3.1.1 Adatok beolvasása	9
3.1.2 Előkészítés	9
3.1.3 Elemzés	9
3.1.4 Publikációkész ábra - 1.	10
3.1.5 Publikációkész ábra - 2.	11
3.1.6 Publikációkész ábra - 3.	12
3.1.7 Publikációkész ábra - 4.	13
3.1.8 Publikációkész ábra - 5.	14
3.2 Eloszlás több csoportra	15
3.2.1 Adatok beolvasása	15
3.2.2 Előkészítés	16
3.2.3 Elemzés	16
3.2.4 Publikációkész ábra - 1.	17
3.2.5 Publikációkész ábra - 2.	18
3.2.6 Publikációkész ábra - 3.	20
3.2.7 Publikációkész ábra - 4.	21
3.2.8 Publikációkész ábra - 5.	23
3.2.9 Publikációkész ábra - 6.	25
4 Dobozdiagram	28
4.1 Dobozdiagram 1 faktorról	28
4.1.1 Adatok beolvasása	28
4.1.2 Előkészítés	28
4.1.3 Elemzés	29
4.1.4 Publikációkész ábra	30
4.2 Dobozdiagram 2 faktorról	31
4.2.1 Adatok beolvasása	31
4.2.2 Előkészítés elemzésre	31

4.2.3	Elemzés	32
4.2.4	Publikációkész ábra - 1.	34
4.2.5	Publikációkész ábra - 2.	35
5	Oszlopdiagram	37
5.1	Egydimenziós oszlopdiagram	37
5.1.1	Adatok beolvasása	37
5.1.2	Előkészítés	37
5.1.3	Elemzés	38
5.1.4	Publikációkész ábra - 1.	39
5.1.5	Publikációkész ábra - 2.	40
5.2	Kétdimenziós oszlopdiagram	42
5.2.1	Adatok beolvasása	42
5.2.2	Előkészítés	42
5.2.3	Elemzés	43
5.2.4	Publikációkész ábra - 1.	45
5.2.5	Publikációkész ábra - 2.	46
5.2.6	Publikációkész ábra - 3.	48
5.2.7	Publikációkész ábra - 4.	49
5.2.8	Publikációkész ábra - 5.	50
5.2.9	Publikációkész ábra - 6.	52
5.2.10	Publikációkész ábra - 7.	54
5.2.11	Publikációkész ábra - 8.	55
5.2.12	Publikációkész ábra - 9.	57
5.2.13	Publikációkész ábra - 10.	58
6	Pontdiagram	61
6.1	Egydimenziós pontdiagram	61
6.1.1	Adatok beolvasása	61
6.1.2	Előkészítés	61
6.1.3	Elemzés	62
6.1.4	Publikációkész ábra - 1.	63
6.1.5	Publikációkész ábra - 2.	64
6.2	Kétdimenziós pontdiagram - csoportosítás nélkül	66
6.2.1	Adatok beolvasása	66
6.2.2	Előkészítés	66
6.2.3	Elemzés	67
6.2.4	Publikációkész ábra - 1.	67
6.2.5	Publikációkész ábra - 2.	68
6.2.6	Publikációkész ábra - 3.	69
6.3	Kétdimenziós pontdiagram - csoportosítással	71
6.3.1	Adatok beolvasása	71
6.3.2	Előkészítés	71

6.3.3	Elemzés	71
6.3.4	Publikációkész ábra - 1.	73
6.3.5	Publikációkész ábra - 2.	74
6.3.6	Publikációkész ábra - 3.	75
7	Átlagábrák	78
7.1	Átlagok egy faktor esetén	78
7.1.1	Adatok beolvasása	78
7.1.2	Előkészítés	78
7.1.3	Elemzés	79
7.1.4	Publikációkész ábra - 1.	80
7.1.5	Publikációkész ábra - 2.	81
7.1.6	Publikációkész ábra - 3.	82
7.1.7	Publikációkész ábra - 4.	84
7.2	Átlagok két faktor esetén	85
7.2.1	Adatok beolvasása	85
7.2.2	Előkészítés	85
7.2.3	Elemzés	86
7.2.4	Publikációkész ábra - 1.	88
7.2.5	Publikációkész ábra - 2.	89
7.2.6	Publikációkész ábra - 3.	91
7.2.7	Publikációkész ábra - 4.	92

Előszó

Jelen jegyzet célja publikációkész ábrák létrehozásának támogatása. Számos eszközben végezhetünk statisztikai elemzéseket, de a minden igénynek eleget tevő ábrák készítésének lehetőségét ma talán a legjobban az R `{ggplot2}` csomagja támogatja.

Jelen dokumentumban a `{ggplot2}` csomag alapján adunk kész receptet az egyes ábrák létrehozására.

1 Bevezetés

A `{ggplot2}` nagyszerű funkcióinak használatához a következő előkészítő lépésekre van szükség:

- Telepítenünk kell a számítógéünkre az R statisztikai programcsomagot, amely letölthető a <https://cran.r-project.org/> címről.
- Telepítenünk kell az RStudio fejlesztői környezetet, amely letölthető a <https://posit.co/download/rstudio-desktop/> címről.
- El kell indítanunk az RStudio programot.
- Létre kell hoznunk egy új projektet az RStudio-ban. Ezt a **File -> New Project** menüpont alatt tehetjük meg. A projektet a **New Directory** opcióval hozhatjuk létre. Válasszuk a **New Project** opciót, majd adjuk meg a projektkönyvtár nevét és helyét. Most feltételezzük, hogy a projekt neve **kesz_abrak** és a helye a **C:/Users/Public/Documents** könyvtárban van. A projekt létrehozása után a RStudio megnyitja a projektet, és a projekt könyvtárában lévő fájlokat jeleníti meg.
- Létre kell hoznunk egy új R szkript fájlt. Ezt a **File -> New File -> R Script** menüpont alatt tehetjük meg. Ezt a fájlt a **kesz_abrak** projekt könyvtárában hozzuk létre. A fájl neve legyen **abrak_publikalasa.R**.
- A **kesz_abrak** projekt könyvtárában hozzunk létre egy új almappát, amelynek a neve **adat** legyen. Ebben az almappában tároljuk majd az ábrák létrehozásához szükséges adatokat. Amennyiben már rendelkezünk az adatokkal, akkor másoljuk be azokat ebbe az almappába. Ha az adataink jamovi **.omv** állományban állnak rendelkezésre, akkor azokat a jamovi program segítségével exportálhatjuk ki ebbe az almappába. Használjuk az SPSS **.sav** formátumot.
- A **kesz_abrak** projekt könyvtárában hozzunk létre egy új almappát, amelynek a neve **output** legyen, majd ebben hozzunk létre egy **kep** nevű almappát. Ebben az almappában tároljuk majd az elkészített ábrákat.

A fenti előkészítés után már készen állunk az ábrák létrehozására. Az első lépésben gondoskodni fogunk a `{ggplot2}` csomag telepítéséről és betöltéséről. A következő lépésben pedig megismerkedünk a `{ggplot2}` csomag nagyszerű funkcióival.

A következő fejezetekben minden R parancsot az `abrak_publikalasa.R` nevű szkriptállományba írjunk, és a `Ctrl+Enter` billentyűkombinációval futtassuk le azokat.

2 Szükséges csomagok

Az első lépésben telepítenünk kell a `{ggplot2}` csomagot. Ezt a következő paranccsal tehetjük meg:

```
# ggplot2 csomag telepítése  
install.packages("ggplot2")
```

Az ábrák létrehozása során minimális adatmanipulációra is szükségünk lehet. Ezt a `{tidyverse}` csomag segítségével végezhetjük el. Telepítsük ezt a csomagot is a következő paranccsal:

```
# a tidyverse csomag telepítése  
install.packages("tidyverse")
```

Az ábrák létrehozását minden esetben az adatok beolvasása előzi meg. Az adatok beolvasására a `{rio}` csomagot használjuk. Telepítsük ezt a csomagot is a következő paranccsal:

```
# a rio csomag telepítése  
install.packages("rio")
```

További csomagok telepítésére is szükség van:

```
# ggk4x csomag a normális eloszlás sűrűségfüggvényének  
# kirajzolásához  
install.packages("ggh4x")  
# jmv csomag a jamovi elemzések futtatásához  
install.packages("jmv")  
# GGally csomag a az oszlopdiagramok helyes  
# megjelenítéséhez  
install.packages("GGally")
```

Amennyiben szeretnénk kipróbálni a jegyzetben lévő R sorokat, használjuk a következő R parancsot, amely létrehozza azt az adatbázist, amelyet a jegyzetben használunk:

```
rio::export(MASS::survey, "adat/survey.xlsx")
```


3 Hisztogram és barátai

3.1 Eloszlás egy csoportra

i Feladat

A `survey.xlsx` 237 egyetemista adatát tartalmazza. A `Height` változó a magasságot tartalmazza cm-ben. Készítsünk hisztogramot és simított hisztogramot a magasság változóra. Vessük össze a normális eloszlás görbéjével.

3.1.1 Adatok beolvasása

Elvégezzük az adatok beolvasását a `survey.xlsx` fájlból. Ha az adatok SPSS `.sav` formátumban állnak rendelkezésre, akkor a beolvasás ugyanígy történik.

```
# Adatok beolvasása
d <- rio::import(file = "adat/survey.xlsx")
```

3.1.2 Előkészítés

Elképzelhető, hogy adataink közvetlenül a beolvasás után még nincsenek elemezhető állapotban. Néhány előkészítő lépést most valóban el kell végeznünk, de ez nem minden adatbázis esetén szükséges. Most sem.

3.1.3 Elemzés

Leíró statisztikai elemzés segítségével megnézzük, hogy milyen értékek ábrázolását fogjuk elvégezni.

```
# leíró statisztikai mutatók a jmv csomag segítségével
jmv::descriptives(data = d, vars = "Height")
#>
#> DESCRIPTIVES
```

```

#>
#> Descriptives
#>
#>                                Height
#>
#>      N                        209
#>   Missing                      28
#>    Mean                     172.3809
#>   Median                     171.0000
#> Standard deviation           9.847528
#>   Minimum                     150.0000
#>   Maximum                     200.0000
#>

```

3.1.4 Publikációkész ábra - 1.

Elkészítjük a publikációkész ábrát. Most hisztogramon jelenítjük meg a változó eloszlását.

```

# ggplot2 csomagok betöltése
library(ggplot2)

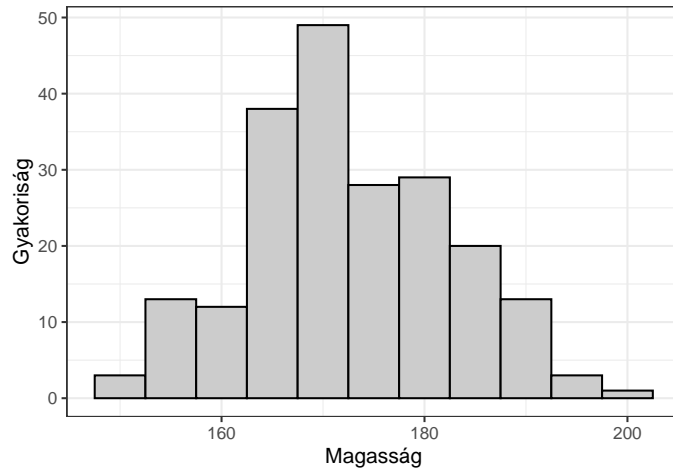
# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Height)) +
  # hisztogram kirajzolása
  geom_histogram(color = "black", fill = "#cccccc", binwidth = 5) +
  # tengelyfeliratok hozzáadása
  labs(x = "Magasság", y = "Gyakoriság") +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/hisztogram_01.png", plot = p1,
        units = "in", width = 6, height = 4, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



3.1.5 Publikációkész ábra - 2.

Elkészítjük a publikációkész ábrát. Most simított hisztogramon jelenítjük meg a változó eloszlását.

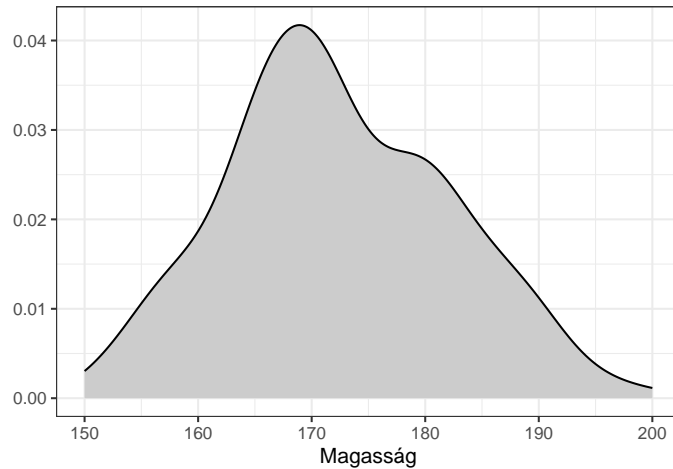
```
# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Height)) +
  # simított hisztogram kirajzolása
  geom_density(color = "black", fill = "#cccccc") +
  # tengelyfeliratok hozzáadása
  labs(x = "Magasság", y=NULL) +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/hisztogram_02.png", plot = p1,
        units = "in", width = 6, height = 4, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre
```



3.1.6 Publikációkész ábra - 3.

Elkészítjük a publikációkész ábrát. Most a hisztogramot és a simított hisztogramot összevetjük.

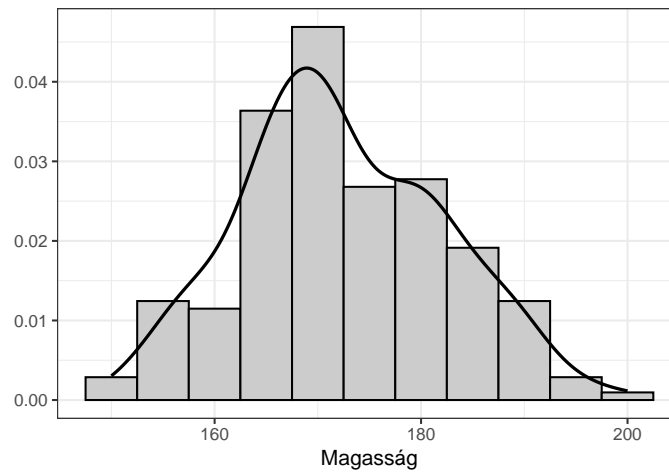
```
# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Height, after_stat(density))) +
  # hisztogram kirajzolása
  geom_histogram(color = "black", fill = "#cccccc", binwidth = 5) +
  # simított hisztogram kirajzolása
  geom_density(color = "black", linewidth=0.8) +
  # tengelyfeliratok hozzáadása
  labs(x = "Magasság", y=NULL) +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/hisztogram_08.png", plot = p1,
        units = "in", width = 6, height = 4, dpi = 300, scale = 0.8)
```

```
p1 # a p1 ábra kirajzolása a képernyőre
```



3.1.7 Publikációkész ábra - 4.

Elkészítjük a publikációkész ábrát. Most a simított hisztogramot összevetjük a normális eloszlás görbéjével (szaggatott vonal).

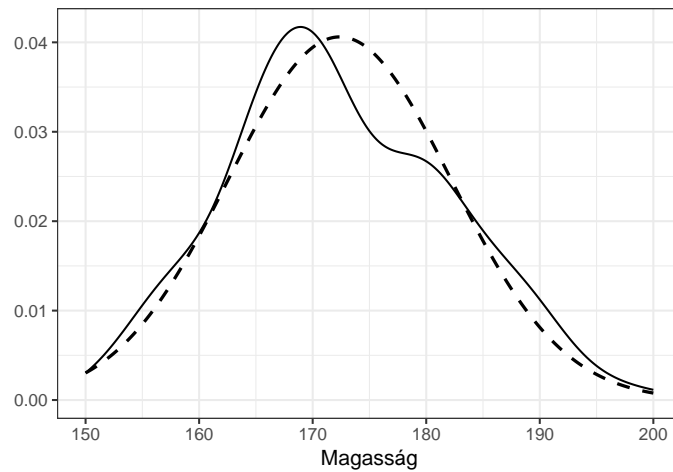
```
# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Height)) +
  # simított hisztogram kirajzolása
  geom_density(color = "black") +
  # normális eloszlás görbéjének kirajzolása
  ggh4x::stat_theodensity(linewidth=0.8, linetype="dashed") +
  # tengelyfeliratok hozzáadása
  labs(x = "Magasság", y=NULL) +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/hisztogram_03.png", plot = p1, units = "in", width = 6, height = 4, dpi = 300)
```

```
p1 # a p1 ábra kirajzolása a képernyőre
```



3.1.8 Publikációkész ábra - 5.

Elkészítjük a publikációkész ábrát. Most a hisztogramot és a simított hisztogramot összevetjük a normális eloszlás görbéjével (szaggatott vonal).

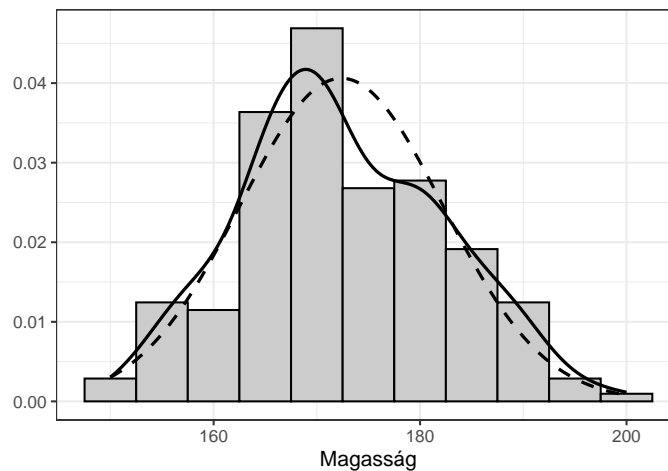
```
# adatok beolvasása
d <- rio::import(file = "adat/survey.xlsx")

# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Height)) +
  # hisztogram kirajzolása
  geom_histogram(aes(y=after_stat(density)), color = "black", fill = "#cccccc", binwidth = 5) +
  # simított hisztogram kirajzolása
  geom_density(color = "black", linewidth=0.8) +
  # normális eloszlás görbéjének kirajzolása
  ggh4x::stat_theodensity(linewidth=0.8, linetype="dashed") +
  # tengelyfeliratok hozzáadása
  labs(x = "Magasság", y=NULL) +
  # fekete-fehér stílus hozzáadása
  theme_bw()
# p1 # a p1 ábra kirajzolása
```

```
# a p1 ábra mentése
ggsave("output/kep/hisztogram_04.png", plot = p1, units = "in", width = 6, height = 4, dpi = 300)

p1 # a p1 ábra kirajzolása a képernyőre
```



3.2 Eloszlás több csoportra

i Feladat

A `survey.xlsx` 237 egyetemista adatát tartalmazza. A `Height` változó a magasságot tartalmazza cm-ben. Készítsünk hisztogramot a magasság változóra a testmozgás gyakorisága (`Exer` változó) szerint csoportosítva.

3.2.1 Adatok beolvasása

Elvégezzük az adatok beolvasását a `survey.xlsx` fájlból. Ha az adatok SPSS `.sav` formátumban állnak rendelkezésre, akkor a beolvasás ugyanígy történik.

```
# Adatok beolvasása
d <- rio::import(file = "adat/survey.xlsx")
```

3.2.2 Előkészítés

Elképzelhető, hogy adataink közvetlenül a beolvasás után még nincsenek elemezhető állapotban. Néhány előkészítő lépést most valóban el kell végeznünk, de ez nem minden adatbázis esetén szükséges:

- a testmozgás gyakoriságát tartalmazó változó (**Exer**) ordinális, így faktorrá kell alakítanunk
- rendezzük az ordinálissá alakított **Exer** változó szintjeit a természetes sorrendbe

```
library(tidyverse)
# faktorrá alakítás
d$Exer <- factor(d$Exer, ordered = T)
# faktor szintjeinek rendezése
d$Exer <- forcats::fct_relevel(d$Exer, "None", "Some", "Freq")
```

3.2.3 Elemzés

Leíró statisztikai elemzés segítségével megnézzük, hogy milyen értékek ábrázolását fogjuk elvégezni.

```
# leíró statisztikai mutatók a jmv csomag segítségével
jmv::descriptives(data = d, vars = "Height", splitBy = "Exer")
#>
#> DESCRIPTIVES
#>
#> Descriptives
#>
#>           Exer      Height
#>
#>      N
#>      None      20
#>      Some      84
#>      Freq     105
#> Missing
#>      None       4
#>      Some      14
#>      Freq      10
#> Mean
#>      None 169.0280
#>      Some 170.3969
#>      Freq 174.6067
#> Median
#>      None 167.0000
#>      Some 168.9500
```



```

#>                               Freq    173.0000
#>   Standard deviation    None    9.470654
#>                               Some    9.471735
#>                               Freq    9.778547
#>   Minimum              None    157.4800
#>                               Some    152.0000
#>                               Freq    150.0000
#>   Maximum              None    190.5000
#>                               Some    193.0400
#>                               Freq    200.0000
#>

```

3.2.4 Publikációkész ábra - 1.

Elkészítjük a publikációkész ábrát. Most hisztogramon jelenítjük meg a változó eloszlását a testmozgás gyakorisága szerint csoportosítva.

```

# ggplot2 csomagok betöltése
library(ggplot2)

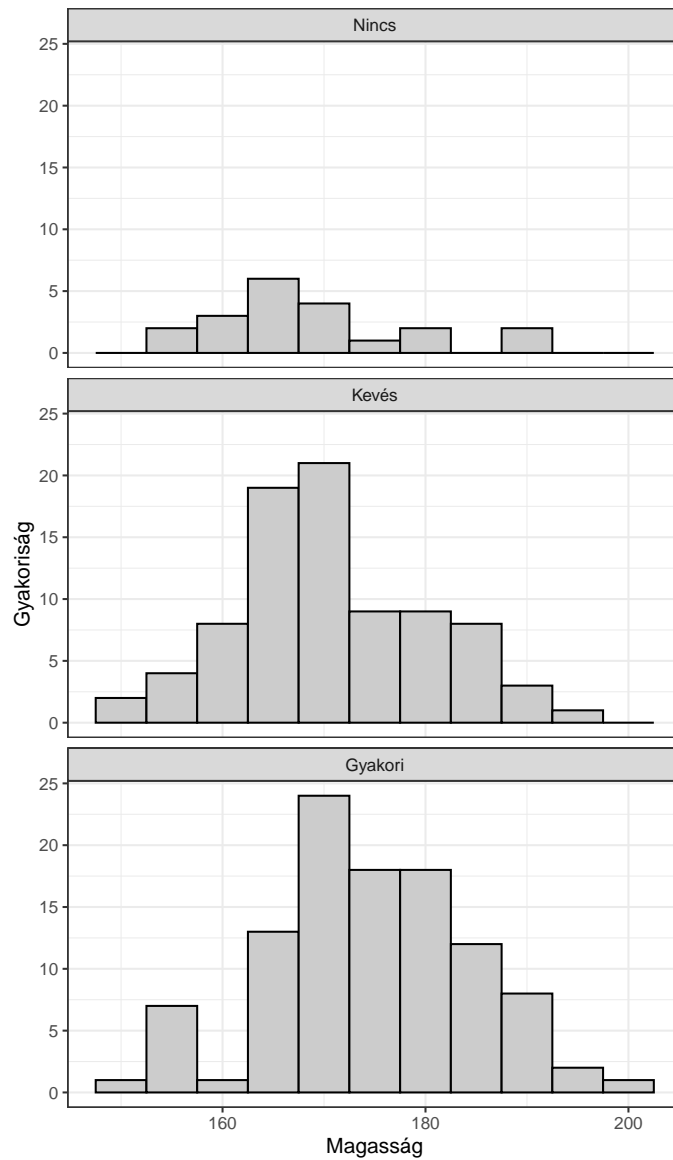
# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Height)) +
  # hisztogram kirajzolása
  geom_histogram(color = "black", fill = "#cccccc", binwidth = 5) +
  # csoportosítás a testmozgás gyakorisága szerint
  facet_wrap(~ Exer, ncol = 1,
             labeller = labeller(Exer = c(None="Nincs", Some="Kevés", Freq="Gyakori")))
  # tengelyfeliratok hozzáadása
  labs(x = "Magasság", y = "Gyakoriság") +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/hisztogram_05.png", plot = p1,
        units = "in", width = 6, height = 10, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



3.2.5 Publikációkész ábra - 2.

Elkészítjük a publikációkész ábrát. Most simított hisztogramon jelenítjük meg a változó eloszlását a testmozgás gyakorisága szerint csoportosítva.

```
# ggplot2 csomagok betöltése
library(ggplot2)
```

```

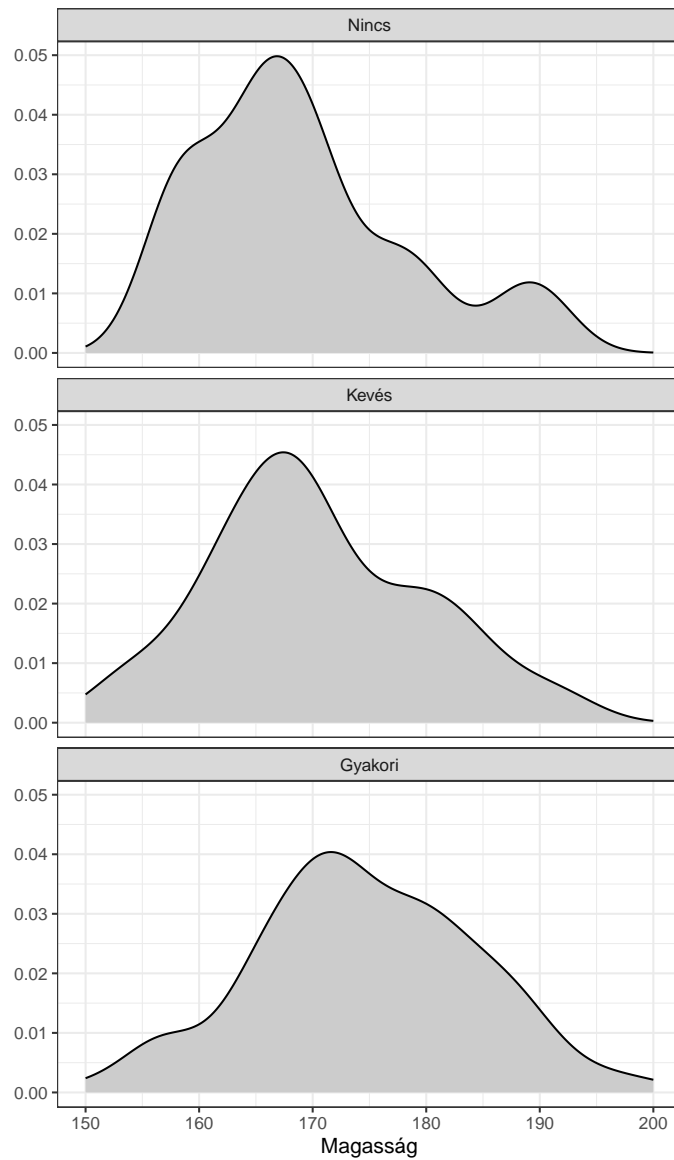
# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Height)) +
  # simított hisztogram kirajzolása
  geom_density(color = "black", fill = "#cccccc") +
  # csoportosítás a testmozgás gyakorisága szerint
  facet_wrap(~ Exer, ncol = 1,
             labeller = labeller(Exer = c(None="Nincs", Some="Kevés", Freq="Gyakori")
  # tengelyfeliratok hozzáadása
  labs(x = "Magasság", y=NULL) +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/hisztogram_04.png", plot = p1,
        units = "in", width = 6, height = 10, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



3.2.6 Publikációkész ábra - 3.

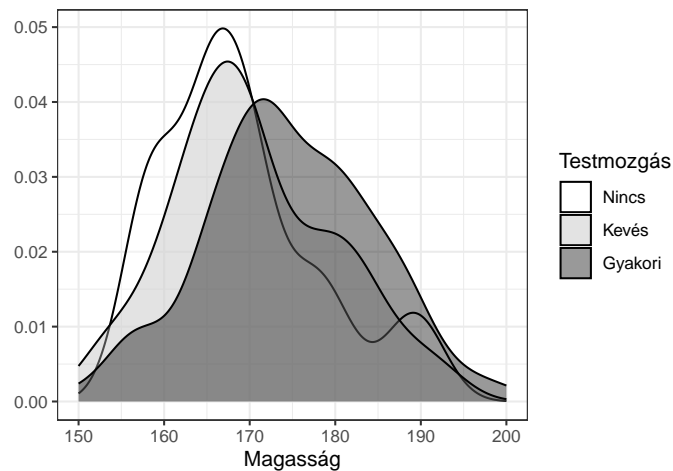
Elkészítjük a publikációkész ábrát. Most az egyes csoportok simított hisztogramjait egyetlen ábrán jelenítjük meg.

```
# ggplot2 csomagok betöltése  
library(ggplot2)
```

```
# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Height, fill=Exer)) +
  # simított hisztogram kirajzolása
  geom_density(color = "black", alpha=0.4) +
  # kitöltőszín beállítása
  scale_fill_grey(start=1, end=0, labels=c(None="Nincs", Some="Kevés", Freq="Gyakori"))
  # tengelyfeliratok hozzáadása
  labs(x = "Magasság", y = NULL, fill="Testmozgás") +
  # fekete-fehér stílus hozzáadása
  theme_bw()
# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/hisztogram_05.png", plot = p1,
        units = "in", width = 6, height = 4, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre
```



3.2.7 Publikációkész ábra - 4.

Elkészítjük a publikációkész ábrát. Most az egyes csoportok hisztogramját és simított hisztogramját egyetlen ábrán jelenítjük meg.

```
# ggplot2 csomagok betöltése
library(ggplot2)
```

```

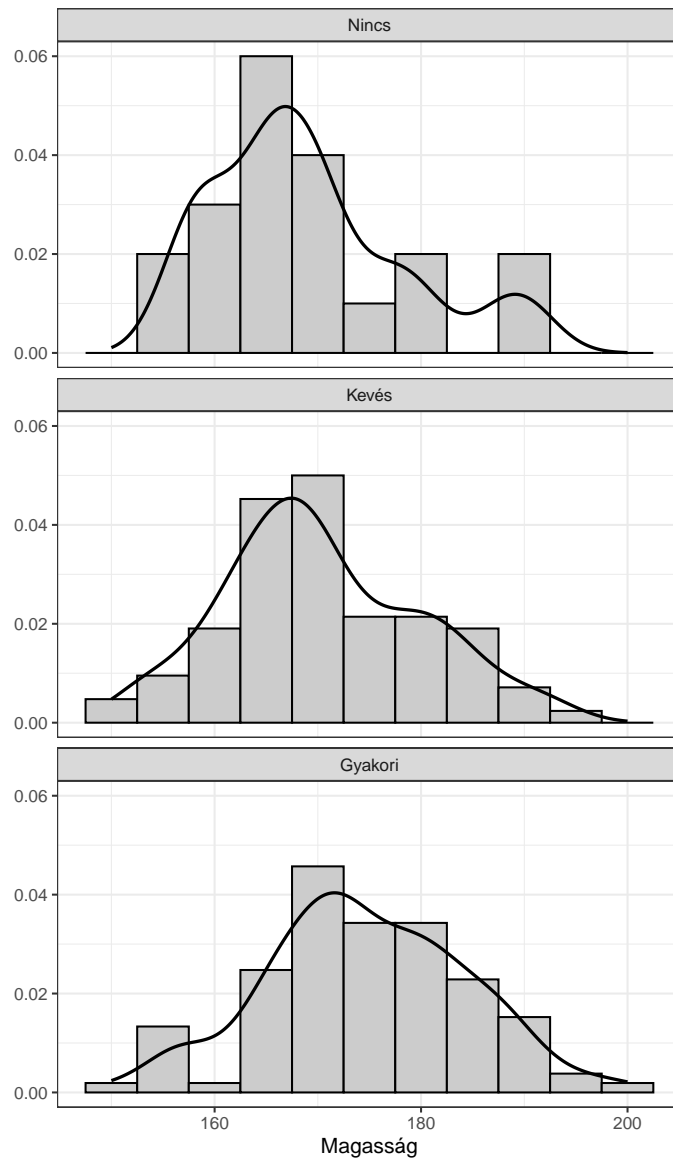
# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Height)) +
  # hisztogram kirajzolása
  geom_histogram(aes(y=after_stat(density)), color = "black", fill = "#cccccc", binw
  # simított hisztogram kirajzolása
  geom_density(color = "black", linewidth=0.8) +
  # csoportosítás a testmozgás gyakorisága szerint
  facet_wrap(~ Exer, ncol = 1,
             labeller = labeller(Exer = c(None="Nincs", Some="Kevés", Freq="Gyakori"
  # tengelyfeliratok hozzáadása
  labs(x = "Magasság", y=NULL) +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/hisztogram_07.png", plot = p1,
        units = "in", width = 6, height = 10, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



3.2.8 Publikációkész ábra - 5.

Elkészítjük a publikációkész ábrát. Most az egyes csoportok hisztogramját és a normális eloszlás gorbéjét is egyetlen ábrán jelenítjük meg.

```
# ggplot2 csomagok betöltése
library(ggplot2)
```

```

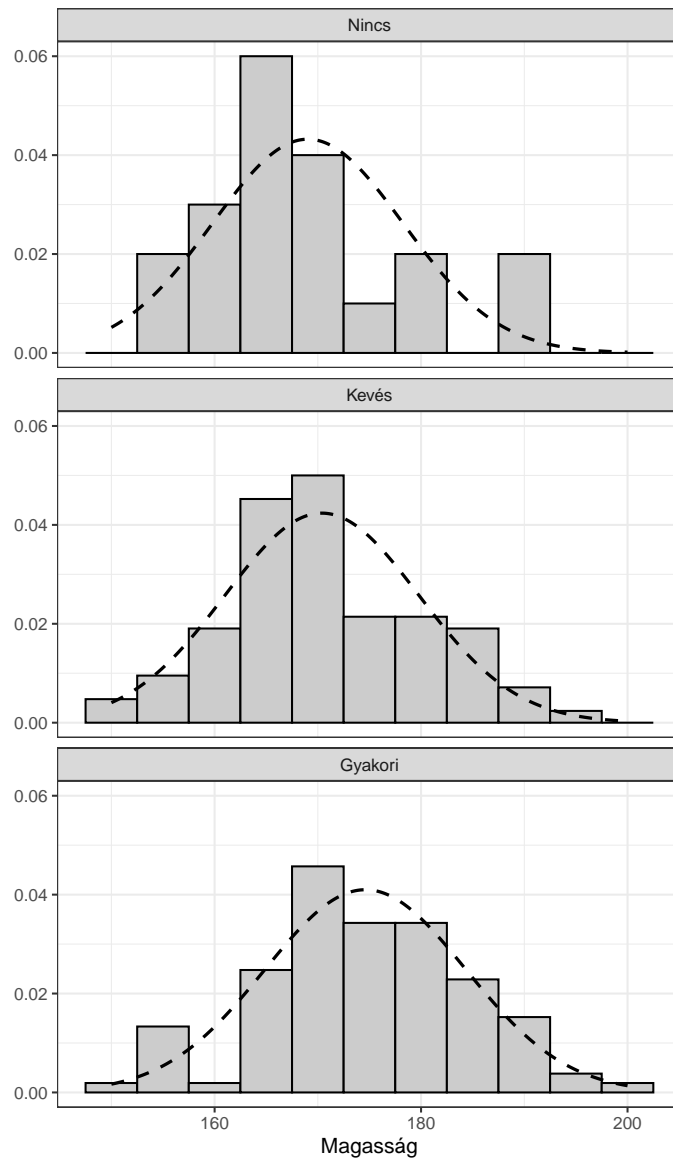
# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Height, y=after_stat(density))) +
  # hisztogram kirajzolása
  geom_histogram(color = "black", fill = "#cccccc", binwidth = 5) +
  # normális eloszlás görbéjének kirajzolása
  ggh4x::stat_theodensity(linewidth=0.8, linetype="dashed") +
  # felosztás a testmozgás gyakorisága szerint
  facet_wrap(~ Exer, ncol = 1,
             labeller = labeller(Exer = c(None="Nincs", Some="Kevés", Freq="Gyakori")
  # tengelyfeliratok hozzáadása
  labs(x = "Magasság", y=NULL) +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/hisztogram_10.png", plot = p1,
       units = "in", width = 6, height = 10, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```

3.2.9 Publikációkész ábra - 6.

Elkészítjük a publikációkész ábrát. Most az egyes csoportok hisztogramját, simított hisztogramját és a normális eloszlás gorbéjét is egyetlen ábrán jelenítjük meg.

```
# ggplot2 csomagok betöltése
library(ggplot2)
```

```

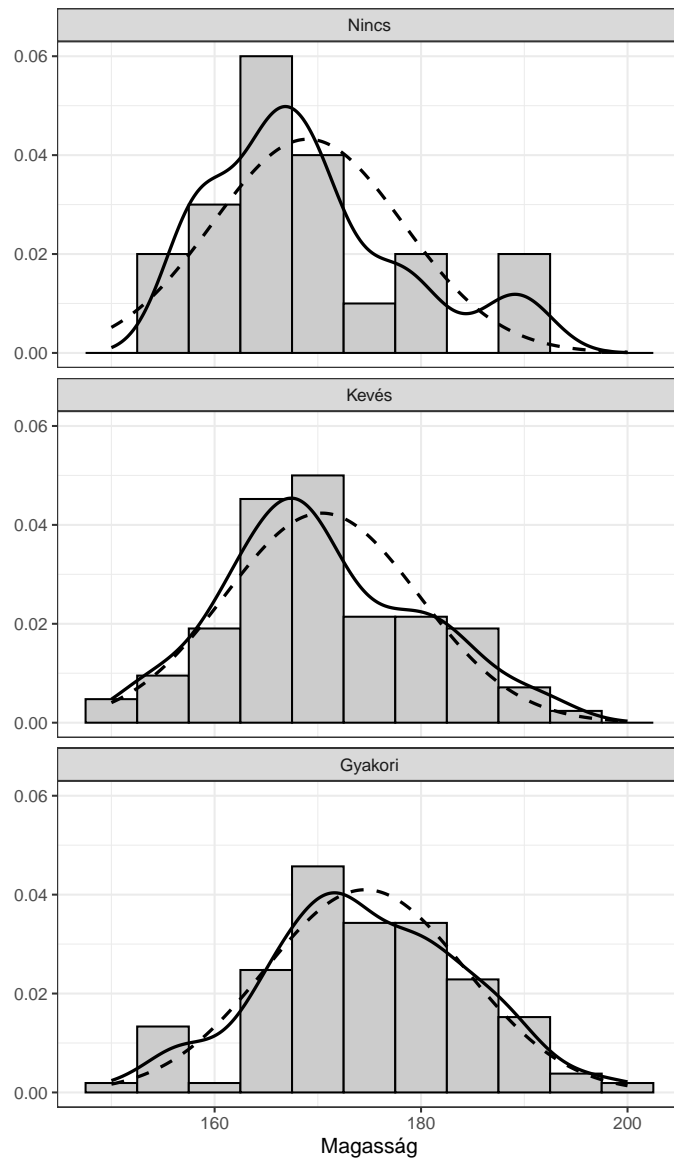
# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Height)) +
  # hisztogram kirajzolása
  geom_histogram(aes(y=after_stat(density)), color = "black", fill = "#cccccc", binw
  # simított hisztogram kirajzolása
  ggh4x::stat_theodensity(linewidth=0.8, linetype="dashed") +
  # simított hisztogram kirajzolása
  geom_density(color = "black", linewidth=0.8) +
  # csoportosítás a testmozgás gyakorisága szerint
  facet_wrap(~ Exer, ncol = 1,
              labeller = labeller(Exer = c(None="Nincs", Some="Kevés", Freq="Gyakori")))
  # tengelyfeliratok hozzáadása
  labs(x = "Magasság", y=NULL) +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/hisztogram_11.png", plot = p1,
        units = "in", width = 6, height = 10, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



4 Dobozdiagram

4.1 Dobozdiagram 1 faktorra

i Feladat

A `survey.xlsx` 237 egyetemista adatát tartalmazza. A `Pulse` változó a pulzust tartalmazza. Készítsünk dobozdiagramot a pulzus változóra a dohányzási szokás figyelembevételével (`Smoke` változó).

4.1.1 Adatok beolvasása

Elvégezzük az adatok beolvasását a `survey.xlsx` fájlból. Ha az adatok SPSS `.sav` formátumban állnak rendelkezésre, akkor a beolvasás ugyanígy történik.

```
# Adatok beolvasása
d <- rio::import(file = "adat/survey.xlsx")
```

4.1.2 Előkészítés

Elképzelhető, hogy adataink közvetlenül a beolvasás után még nincsenek elemezhető állapotban. Néhány előkészítő lépést most valóban el kell végeznünk, de ez nem minden adatbázis esetén szükséges:

- a dohányzási szokásokat tartalmazó változó (`Smoke`) ordinális, így faktorrá kell alakítanunk
- rendezzük az ordinálissá alakított `Smoke` változó szintjeit a természetes sorrendbe.

```
library(tidyverse)
# faktorrá alakítás
d$Smoke <- factor(d$Smoke, ordered = T)
# faktor szintjeinek rendezése
d$Smoke <- forcats::fct_relevel(d$Smoke, "Never", "Occas", "Regul",
```

```
"Heavy")
```

4.1.3 Elemzés

Leíró statisztikai elemzés segítségével megnézzük, hogy milyen értékek ábrázolását fogjuk elvégezni.

```
# leíró statisztikai mutatók a jmv csomag segítségével
jmv::descriptives(data = d, vars = "Pulse", split = "Smoke")
#>
#> DESCRIPTIVES
#>
#> Descriptives
#>
#>
#>      Smoke      Pulse
#>
#> N      Never      152
#>      Occas       16
#>      Regul       16
#>      Heavy        7
#> Missing Never      37
#>      Occas        3
#>      Regul        1
#>      Heavy        4
#> Mean      Never  73.99342
#>      Occas  73.93750
#>      Regul  73.68750
#>      Heavy  78.28571
#> Median    Never  72.00000
#>      Occas  76.00000
#>      Regul  69.00000
#>      Heavy  80.00000
#> Standard deviation Never  11.71035
#>      Occas  9.226547
#>      Regul  14.42553
#>      Heavy  11.70063
#> Minimum    Never  35.00000
#>      Occas  60.00000
#>      Regul  50.00000
#>      Heavy  65.00000
```

```
#>      Maximum      Never      104.0000
#>      Occas      87.00000
#>      Regul      104.0000
#>      Heavy      100.0000
#>
```

4.1.4 Publikációkész ábra

Elkészítjük a publikációkész ábrát.

```
# ggplot2 csomagok betöltése
library(ggplot2)

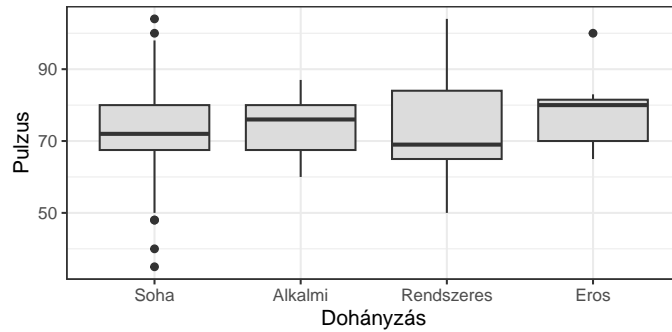
# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, y=Pulse)) +
  # dobozdiagram kirajzolása
  geom_boxplot(fill="#DCDCDC") +
  # az x tengely beállítása
  scale_x_discrete(na.translate = FALSE,
                   labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +

  # feliratok hozzáadása
  labs(x = "Dohányzás", y = "Pulzus") +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/dobozdiagram_01.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre
```



4.2 Dobozdiagram 2 faktorral

i Feladat

A `survey.xlsx` 237 egyetemista adatát tartalmazza. A `Pulse` változó a pulzust tartalmazza. Készítsünk dobozdiagramot a pulzus változóra a dohányzási szokás (`Smoke` változó) és a nem (`Sex`) figyelembevételével.

4.2.1 Adatok beolvasása

Elvégezzük az adatok beolvasását a `survey.xlsx` fájlból. Ha az adatok SPSS `.sav` formátumban állnak rendelkezésre, akkor a beolvasás ugyanígy történik.

```
# Adatok beolvasása
d <- rio::import(file = "adat/survey.xlsx")
```

4.2.2 Előkészítés elemzésre

Elképzelhető, hogy adataink közvetlenül a beolvasás után még nincsenek elemezhető állapotban. Néhány előkészítő lépést most valóban el kell végeznünk, de ez nem minden adatbázis esetén szükséges:

- a nem változó (`Sex`) nominális, így faktorrá kell alakítanunk,
- a dohányzási szokásokat tartalmazó változó (`Smoke`) ordinális, így faktorrá kell alakítanunk,
- rendezzük az ordinálissá alakított `Smoke` változó szintjeit a természetes sorrendbe,
- azokkal a személyekkel nem szeretnénk elemzést végezni, akiknek nincs megadva a nemük, így ezeket a sorokat kiszűrjük az adatbázisból.

```

library(tidyverse)
# faktorrá alakítás
d$Sex <- factor(d$Sex)
# faktorrá alakítás
d$Smoke <- factor(d$Smoke)
# faktor szintjeinek rendezése
d$Smoke <- forcats::fct_relevel(d$Smoke, "Never", "Occas", "Regul",
  "Heavy")
# kiszűrjük azokat a személyeket, akiknek nincs megadva a
# nemük
d <- d |>
  dplyr::filter(!is.na(Sex))

```

4.2.3 Elemzés

Leíró statisztikai elemzés segítségével megnézzük, hogy milyen értékek ábrázolását fogjuk elvégezni.

```

# leíró statisztikai mutatók a jmv csomag segítségével
library(jmv)
jmv::descriptives(data = d, vars = "Pulse", split = c("Smoke",
  "Sex"))

```

	Smoke	Sex	Pulse
N	Never	Female	7...
		Male	7...
	Occas	Female	...
		Male	...
	Regul	Female	...
		Male	1...
	Heavy	Female	...
		Male	...
Missing	Never	Female	2...
		Male	1...
	Occas	Female	...

#>			Male	...
#>		Regul	Female	...
#>			Male	...
#>		Heavy	Female	...
#>			Male	...
#>	Mean	Never	Female	75.6075...
#>			Male	72.2361...
#>		Occas	Female	73.3750...
#>			Male	74.5000...
#>		Regul	Female	69.2500...
#>			Male	75.1666...
#>		Heavy	Female	75.0000...
#>			Male	82.6666...
#>	Median	Never	Female	75.0000...
#>			Male	72.0000...
#>		Occas	Female	78.0000...
#>			Male	73.5000...
#>		Regul	Female	67.5000...
#>			Male	72.0000...
#>		Heavy	Female	76.0000...
#>			Male	83.0000...
#>	Standard deviation	Never	Female	11.4078...
#>			Male	11.9429...
#>		Occas	Female	11.2369...
#>			Male	7.44503...
#>		Regul	Female	17.3853...
#>			Male	13.8487...
#>		Heavy	Female	6.00000...
#>			Male	17.5023...
#>	Minimum	Never	Female	40.0000...
#>			Male	35.0000...
#>		Occas	Female	60.0000...
#>			Male	66.0000...
#>		Regul	Female	50.0000...
#>			Male	54.0000...
#>		Heavy	Female	68.0000...
#>			Male	65.0000...
#>	Maximum	Never	Female	104.000...
#>			Male	97.0000...
#>		Occas	Female	87.0000...
#>			Male	87.0000...

```

#>               Regul    Female    92.0000...
#>               Male      104.000...
#>               Heavy    Female    80.0000...
#>               Male      100.000...
#>

```

4.2.4 Publikációkész ábra - 1.

Elkészítjük a publikációkész ábrát. Itt azt a megoldást választottuk, hogy a nem (Sex) szerint kettébontjuk az ábrát.

```

# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, y=Pulse)) +
  # dobozdiagram kirajzolása
  geom_boxplot(fill="#DCDCDC") +
  # az x tengely beállítása
  scale_x_discrete(na.translate = FALSE,
                   labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +

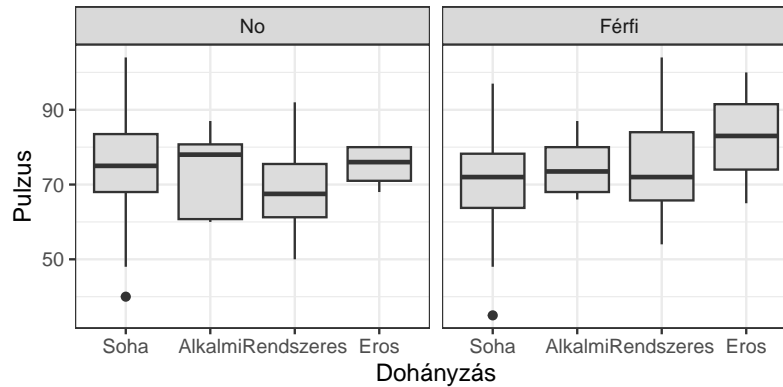
  # feliratok hozzáadása
  labs(x = "Dohányzás", y = "Pulzus") +
  # panelekre osztjuk a rajzterületet a nem változó alapján
  facet_wrap(~ Sex, nrow = 1,
             labeller = labeller(Sex=c(Female="Nő", Male="Férfi")) +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/dobozdiagram_02.png", plot = p1,
        units = "in", width = 7, height = 7*0.5, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



4.2.5 Publikációkész ábra - 2.

Elkészítjük a publikációkész ábrát. Itt egyetlen ábrán jelenítjük meg a dohányzási szokásokat és a nemeket.

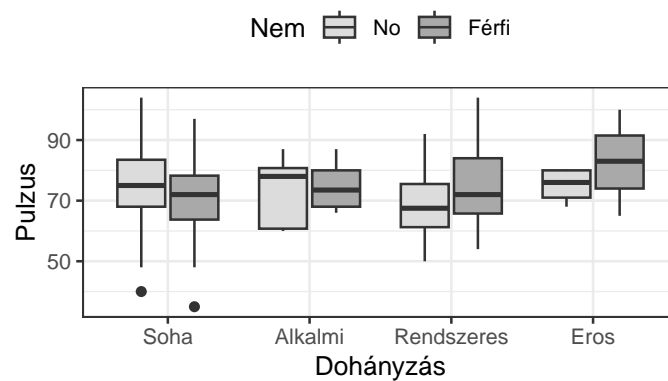
```
# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, y=Pulse, fill=Sex)) +
  # dobozdiagram kirajzolása
  geom_boxplot() +
  # az x tengely beállítása
  scale_x_discrete(na.translate = FALSE,
                   labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +
  # a kitöltés (fill) paraméterek beállítása
  scale_fill_manual(values = c("#DCDCDC", "#A9A9A9"),
                   label=c(Female="Nő", Male="Férfi")) +
  # feliratok hozzáadása
  labs(x = "Dohányzás", y = "Pulzus", fill="Nem") +
  # fekete-fehér stílus hozzáadása
  theme_bw() +
  # a jelmagyarázat pozíciójának beállítása
  theme(legend.position = "top")

# p1 # a p1 ábra kirajzolása
```

```
# a p1 ábra mentése
ggsave("output/kep/dobozdiagram_03.png", plot = p1,
       units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)
```

```
p1 # a p1 ábra kirajzolása a képernyőre
```



5 Oszlopdiaagram

5.1 Egydimenziós oszlopdiaagram

Feladat

A `survey.xlsx` 237 egyetemista adatát tartalmazza. A `Smoke` változó a dohányzásra vonatkozó kérdésre adott válaszokat tartalmazza. Rajzoljuk meg a dohányzási szokás eloszlását. Jellemezzük az eloszlást leíró, számszerű értékekkel is.

5.1.1 Adatok beolvasása

Elvégezzük az adatok beolvasását a `survey.xlsx` fájlból. Ha az adatok SPSS `.sav` formátumban állnak rendelkezésre, akkor a beolvasás ugyanígy történik.

```
# Adatok beolvasása
d <- rio::import(file = "adat/survey.xlsx")
```

5.1.2 Előkészítés

Elképzelhető, hogy adataink közvetlenül a beolvasás után még nincsenek elemezhető állapotban. Néhány előkészítő lépést most valóban el kell végeznünk, de ez nem minden adatbázis esetén szükséges:

- a dohányzási szokásokat tartalmazó változó (`Smoke`) ordinális, így faktorrá kell alakítanunk,
- rendezzük az ordinálissá alakított `Smoke` változó szintjeit a természetes sorrendbe.

```
library(tidyverse)
# faktorrá alakítás
d$Smoke <- factor(d$Smoke, ordered = T)
# faktor szintjeinek rendezése
d$Smoke <- forcats::fct_relevel(d$Smoke, "Never", "Occas", "Regul",
```

```
"Heavy")
```

5.1.3 Elemzés

Leíró statisztikai elemzés segítségével megnézzük, hogy milyen értékek ábrázolását fogjuk elvégezni.

```
# leíró statisztikai mutatók a jmv csomag segítségével
jmv::descriptives(data = d, vars = "Smoke", freq = TRUE)
#>
#> DESCRIPTIVES
#>
#> Descriptives
#>
#>
#> Smoke
#>
#> N 236
#> Missing 1
#> Mean
#> Median
#> Standard deviation
#> Minimum
#> Maximum
#>
#>
#>
#> FREQUENCIES
#>
#> Frequencies of Smoke
#>
#> Smoke Counts % of Total Cumulative %
#>
#> Never 189 80.08475 80.08475
#> Occas 19 8.05085 88.13559
#> Regul 17 7.20339 95.33898
#> Heavy 11 4.66102 100.00000
#>
```

5.1.4 Publikációkész ábra - 1.

Elkészítjük a publikációkész ábrát. Most olyan megjelenítést választottunk, amelyben **az** egyes kategóriák abszolút gyakoriságát jelenítjük meg.

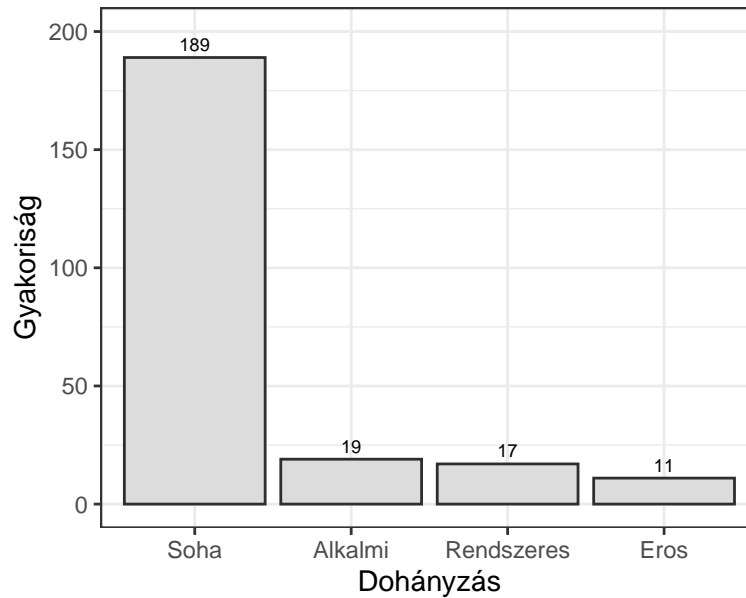
```
# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke)) +
  # oszlopdiagram kirajzolása
  geom_bar(fill="#dcdcdc", col="#2e2e2e") +
  # az x tengely beállítása
  scale_x_discrete(na.translate = FALSE,
                    labels=c(Never="Soha",
                             Occas="Alkalmi",
                             Regul="Rendszeres",
                             Heavy="Erős")) +
  # gyakoriságok megjelenítése
  geom_text(stat = "count", size=2.6,
            aes(label = after_stat(count)), vjust = -0.5) +
  # az y tengely láthatóságának beállítása
  coord_cartesian(ylim = c(0, 200)) +
  # feliratok hozzáadása
  labs(x = "Dohányzás", y = "Gyakoriság") +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/oszlopdiagram_01.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre
```



5.1.5 Publikációkész ábra - 2.

Újabb ábrát készítünk. Most olyan megjelenítést választottunk, amelyben az egyes kategóriák százalékos relatív gyakoriságát jelenítjük meg.

```
# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, y=after_stat(count/sum(count)))) +
  # oszlopdiagram kirajzolása
  geom_bar(fill="#dcdcdc", col="#2e2e2e") +
  # az x tengely beállítása
  scale_x_discrete(na.translate = FALSE,
                   labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +
  # százalékos gyakoriságok megjelenítése
  geom_text(stat = "count", vjust = -0.5, size=2.6,
           aes(label =
               scales::percent(after_stat(count/sum(count)),
                               decimal.mark = ","),
```



```

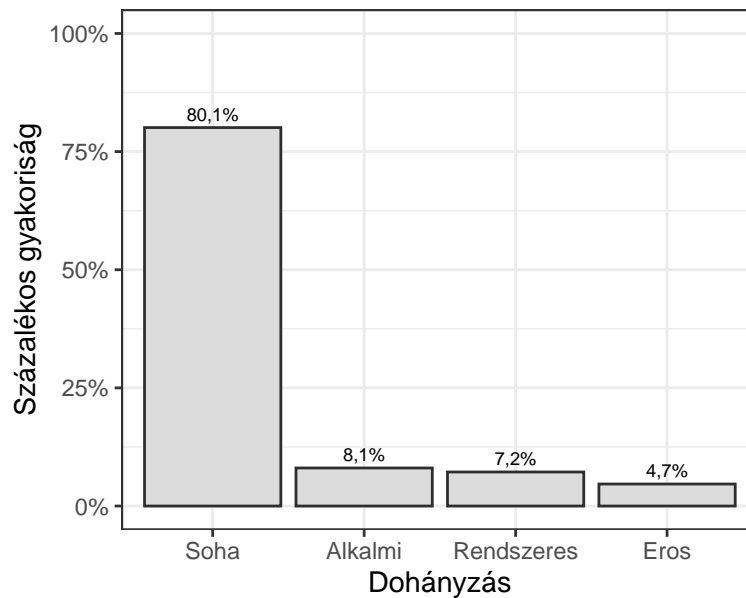
accuracy = 0.1))) +
# feliratok hozzáadása
labs(x = "Dohányzás", y = "Százalékos gyakoriság") +
# y tengely formázása
scale_y_continuous(labels = scales::percent) +
# az y tengely láthatóságának beállítása
coord_cartesian(ylim = c(0, 1)) +
# fekete-fehér stílus hozzáadása
theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/oszlopdiagram_01.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



5.2 Kétdimenziós oszlopdiagram

i Feladat

A `survey.xlsx` 237 egyetemista adatát tartalmazza. A `Smoke` változó a dohányzásra vonatkozó kérdésre adott válaszokat tartalmazza. A `Sex` változó a nemet tartalmazza. Rajzoljuk meg a dohányzási szokás és a nem kapcsolatát. Jellemezzük az eloszlást leíró, számszerű értékekkel is.

5.2.1 Adatok beolvasása

Elvégezzük az adatok beolvasását a `survey.xlsx` fájlból. Ha az adatok SPSS `.sav` formátumban állnak rendelkezésre, akkor a beolvasás ugyanígy történik.

```
# Adatok beolvasása
d <- rio::import(file = "adat/survey.xlsx")
```

5.2.2 Előkészítés

Elképzelhető, hogy adataink közvetlenül a beolvasás után még nincsenek elemezhető állapotban. Néhány előkészítő lépést most valóban el kell végeznünk, de ez nem minden adatbázis esetén szükséges:

- a nem változó (`Sex`) nominális, így faktorrá kell alakítanunk,
- a dohányzási szokásokat tartalmazó változó (`Smoke`) ordinális, így faktorrá kell alakítanunk,
- rendezzük az ordinálissá alakított `Smoke` változó szintjeit a természetes sorrendbe,
- azokkal a személyekkel nem szeretnénk elemzést végezni, akiknek nincs megadva a nemük, vagy nem ismerjük a dohányzási szokásukat, így ezeket a sorokat kiszűrjük az adatbázisból.

```
library(tidyverse)
# faktorrá alakítás
d$Sex <- factor(d$Sex)
# faktorrá alakítás
d$Smoke <- factor(d$Smoke)
# faktor szintjeinek rendezése
d$Smoke <- forcats::fct_relevel(d$Smoke, "Never", "Occas", "Regul",
                                "Heavy")
```

```
# kiszűrjük azokat a személyeket, akiknek nincs megadva a
# nemük vagy a dohányzási szokásuk
d <- d |>
  dplyr::filter(!is.na(Sex) & !is.na(Smoke))
```

5.2.3 Elemzés

Leíró statisztikai elemzés segítségével megnézzük, hogy milyen értékek ábrázolását fogjuk elvégezni.

```
# leíró statisztikai mutatók a jmv csomag segítségével
jmv::descriptives(data = d, vars = "Smoke", splitBy = "Sex",
  freq = TRUE)
#>
#> DESCRIPTIVES
#>
#> Descriptives
#>
#>           Sex      Smoke
#>
#> N           Female    118
#>           Male       117
#> Missing      Female     0
#>           Male       0
#> Mean          Female
#>           Male
#> Median         Female
#>           Male
#> Standard deviation Female
#>           Male
#> Minimum         Female
#>           Male
#> Maximum         Female
#>           Male
#>
#>
#> FREQUENCIES
#>
#> Frequencies of Smoke ...
```

```

#>
#>   Smoke   Sex   Counts   % of Total   Cumul...
#>
#>   Never   Female    99    42.12766    4...
#>           Male     89    37.87234    8...
#>   Occas   Female     9     3.82979    8...
#>           Male    10     4.25532    8...
#>   Regul   Female     5     2.12766    9...
#>           Male    12     5.10638    9...
#>   Heavy   Female     5     2.12766    9...
#>           Male     6     2.55319   10...
#>
#>   ...

```

```

# leíró statisztikai mutatók a jmv csomag segítségével
jmv::contTables(data = d, rows = "Sex", cols = "Smoke", pcRow = T,
  pcCol = T, pcTot = T)

```

```

#>
#> CONTINGENCY TABLES
#>
#> Contingency Tables
#>
#>   Sex
#>   ...
#>   Never   Occas ...
#>   ...
#>   Female   Observed    99    ...
#>           % within row  83.89831  7.62...
#>           % within column 52.65957 47.36...
#>           % of total    42.12766  3.82...
#>
#>   Male     Observed    89    ...
#>           % within row  76.06838  8.54...
#>           % within column 47.34043 52.63...
#>           % of total    37.87234  4.25...
#>
#>   Total    Observed   188    ...
#>           % within row  80.00000  8.08...
#>           % within column 100.00000 100.00...
#>           % of total    80.00000  8.08...
#>
#>   ...
#>
#>
#>   ^ Tests

```

```

#>
#>           Value      df      p
#>
#>      2      3.553618      3      0.3138744
#>      N          235
#>

```

5.2.4 Publikációkész ábra - 1.

Elkészítjük az ábrát. Most olyan megjelenítést választottunk, amelyben **az egyes kategóriák abszolút gyakoriságát jelenítjük meg**. Az x tengelyen jelenítjük meg a dohányzási szokásokat, a kitöltés paraméterrel a nemeket különböztetjük meg, és az azonos dohányzási kategóriához tartozó gyakoriságokat **egymásra halmozzuk**.

```

# ggplot2 csomagok betöltése
library(ggplot2)

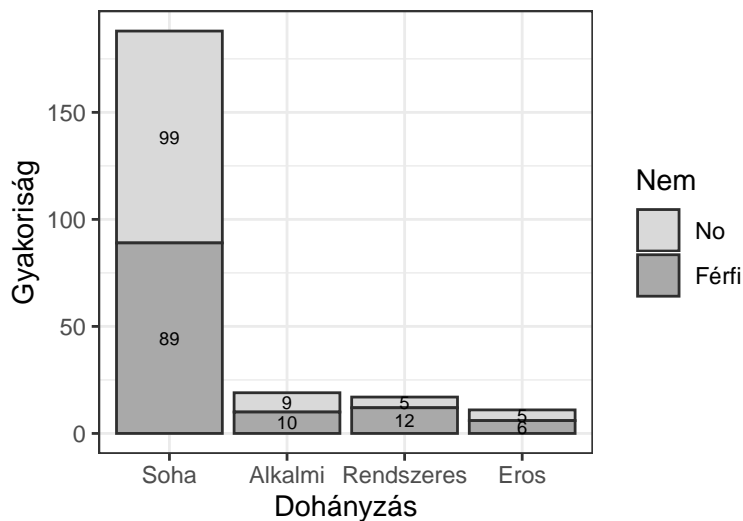
# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, fill=Sex)) +
  # oszlopdiagram kirajzolása
  geom_bar(col="#2e2e2e", position="stack") +
  # az x tengely beállítása
  scale_x_discrete(na.translate = FALSE,
                   labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +
  # a kitöltés (fill) paraméterek beállítása
  scale_fill_manual(values = c("#D9D9D9", "#A9A9A9"),
                   label=c(Female="Nő", Male="Férfi")) +
  # gyakoriságok megjelenítése
  geom_text(stat = "count", size=2.6,
           aes(label = after_stat(count)),
           position = position_stack(vjust = 0.5)) +
  # feliratok hozzáadása
  labs(x = "Dohányzás", y = "Gyakoriság", fill="Nem") +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

```

```
# a p1 ábra mentése
ggsave("output/kep/oszlopdiagram_03.png", plot = p1,
       units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)
```

```
p1 # a p1 ábra kirajzolása a képernyőre
```



5.2.5 Publikációkész ábra - 2.

Újabb publikációkész ábrát készítünk. Most is olyan megjelenítést választottunk, amelyben **az egyes kategóriák abszolút gyakoriságát jelenítjük meg**. Az x tengelyen azonban a nemeket jelenítjük meg, a kitöltés paraméterrel a dohányzási szokásokat különböztetjük meg, és az azonos nemhez tartozó gyakoriságokat **egymásra halmozzuk**.

```
# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Sex, fill=Smoke)) +
  # oszlopdiagram kirajzolása
  geom_bar(col="#2e2e2e", position="stack") +
  # az x tengely beállítása
  scale_x_discrete(labels=c(Female="Nő", Male="Férfi")) +
  # a kitöltés (fill) paraméterek beállítása
  scale_fill_manual(values = c("#E9E9E9", "#C9C9C9", "#A9A9A9", "#898989"),
```

```

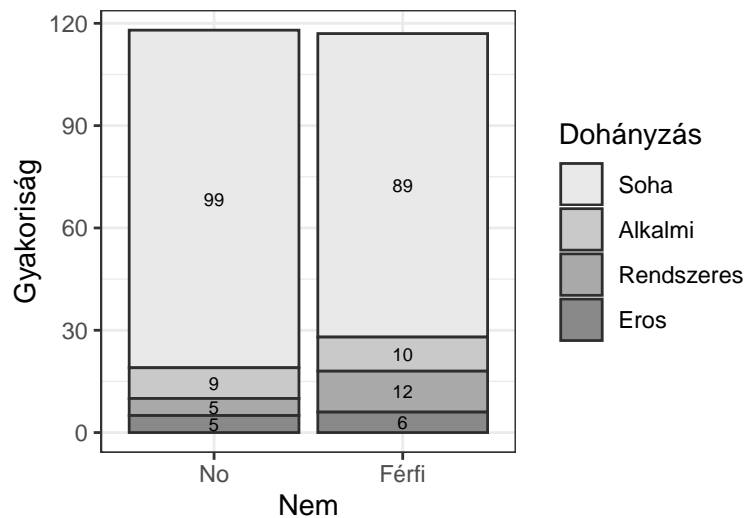
na.translate = FALSE,
labels=c(Never="Soha",
        Occas="Alkalmi",
        Regul="Rendszeres",
        Heavy="Erős")) +
# gyakoriságok megjelenítése
geom_text(stat = "count", size=2.6,
          aes(label = after_stat(count)),
          position = position_stack(vjust = 0.5)) +
# feliratok hozzáadása
labs(x = "Nem", y = "Gyakoriság", fill="Dohányzás") +
# fekete-fehér stílus hozzáadása
theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/oszlopdiagram_04.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



5.2.6 Publikációkész ábra - 3.

Elkészítjük a publikációkész ábrát. Most olyan megjelenítést választottunk, amelyben **az egyes kategóriák abszolút gyakoriságát jelenítjük meg**. Az x tengelyen jelenítjük meg a dohányzási szokásokat, a kitöltés paraméterrel a nemeket különböztetjük meg és **egy más melletti oszlopokba írjuk őket**.

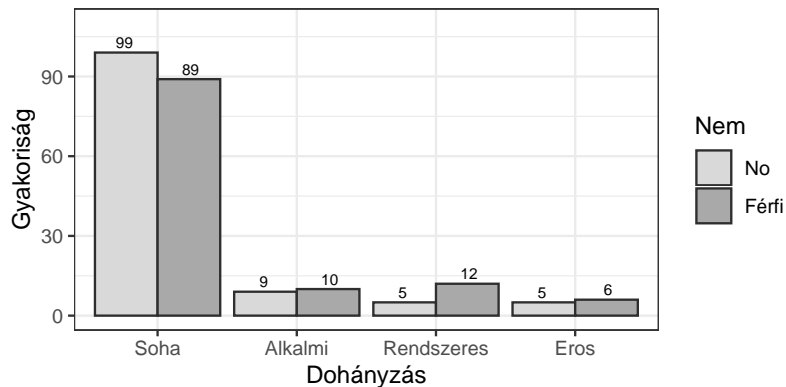
```
# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, fill=Sex)) +
  # oszlopdiagram kirajzolása
  geom_bar(col="#2e2e2e", position="dodge") +
  # az x tengely beállítása
  scale_x_discrete(na.translate = FALSE,
                   labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +
  # a kitöltés (fill) paraméterek beállítása
  scale_fill_manual(values = c("#D9D9D9", "#A9A9A9"),
                   label=c(Female="Nő", Male="Férfi")) +
  # gyakoriságok megjelenítése
  geom_text(stat = "count", size=2.6,
           aes(label = after_stat(count)),
           position=position_dodge(width=0.95), vjust=-0.4) +
  # feliratok hozzáadása
  labs(x = "Dohányzás", y = "Gyakoriság", fill="Nem") +
  # y tengely láthatóságának beállítása
  coord_cartesian(ylim = c(0, 110)) +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/oszlopdiagram_03.png", plot = p1,
       units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre
```

5.2.7 Publikációkész ábra - 4.

Elkészítjük a publikációkész ábrát. Most olyan megjelenítést választottunk, amelyben **az egyes kategóriák abszolút gyakoriságát jelenítjük meg**. Az x tengelyen jelenítjük meg a nemeket, a kitöltés paraméterrel a dohányzási szokásokat különböztetjük meg és **egymás melletti oszlopokba** írjuk őket.

```
# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Sex, fill=Smoke)) +
  # oszlopdiagram kirajzolása
  geom_bar(col="#2e2e2e", position="dodge") +
  # az x tengely beállítása
  scale_x_discrete(labels=c(Female="Nő", Male="Férfi")) +
  # a kitöltés (fill) paraméterek beállítása
  scale_fill_manual(values = c("#E9E9E9", "#C9C9C9", "#A9A9A9", "#898989"),
    na.translate = FALSE,
    labels=c(Never="Soha",
      Occas="Alkalmi",
      Regul="Rendszeres",
      Heavy="Erős")) +

  # gyakoriságok megjelenítése
  geom_text(stat = "count", size=2.6,
    aes(label = after_stat(count)),
    position=position_dodge(width=0.95), vjust=-0.4) +

  # y tengely láthatóságának beállítása
  coord_cartesian(ylim = c(0, 110)) +
```

```

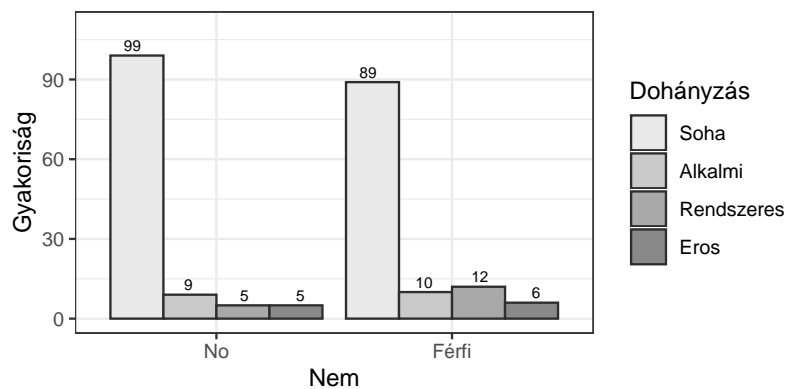
# feliratok hozzáadása
labs(x = "Nem", y = "Gyakoriság", fill="Dohányzás") +
# fekete-fehér stílus hozzáadása
theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/oszlopdiagram_04.png", plot = p1,
       units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



5.2.8 Publikációkész ábra - 5.

Elkészítjük a publikációkész ábrát. Most olyan megjelenítést választottunk, amelyben **az egyes kategóriák százalékos relatív gyakoriságát jelenítjük meg a teljes mintára vonatkoztatva (a megjelenített százalékos értékek összege 100%)**. Az x tengelyen jelenítjük meg a dohányzási szokásokat, a kitöltés paraméterrel a nemeket különböztetjük meg, és az azonos dohányzási kategóriába tartozó százalékos gyakoriságokat **egymásra halmozuk**.

```

# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, fill=Sex, y=after_stat(count/sum(count)))) +
# oszlopdiagram kirajzolása

```

```

geom_bar(col="#2e2e2e", position="stack") +
# az x tengely beállítása
scale_x_discrete(na.translate = FALSE,
                  labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +
# a kitöltés (fill) paraméterek beállítása
scale_fill_manual(values = c("#D9D9D9", "#A9A9A9"),
                  label=c(Female="Nő", Male="Férfi")) +
# százalékos gyakoriságok megjelenítése
geom_text(stat = "count", size=2.6,
          position = position_stack(vjust = 0.5),
          aes(label =
               scales::percent(after_stat(count/sum(count)),
                                decimal.mark = ",",
                                accuracy = 0.1))) +

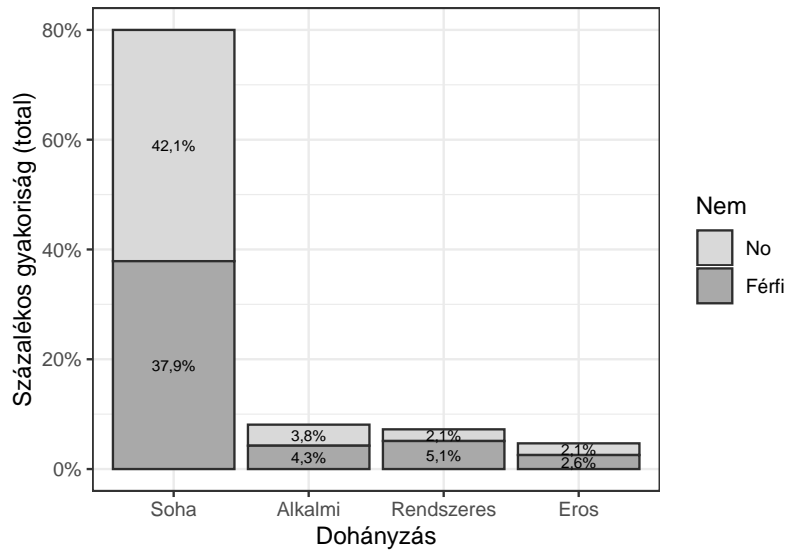
# feliratok hozzáadása
labs(x = "Dohányzás", y = "Százalékos gyakoriság (total)", fill="Nem") +
# y tengely formázása
scale_y_continuous(labels = scales::percent) +
# fekete-fehér stílus hozzáadása
theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/oszlopdiagram_03.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



5.2.9 Publikációkész ábra - 6.

Elkészítjük a publikációkész ábrát. Most olyan megjelenítést választottunk, amelyben **az egyes kategóriák százalékos relatív gyakoriságát jelenítjük meg a teljes mintára vonatkoztatva (a megjelenített százalékos értékek összege 100%)**. Az x tengelyen jelenítjük meg a nemeket, a kitöltés paraméterrel a dohányzási szokásokat különböztetjük meg, és az azonos nemhez tartozó százalékos gyakoriságokat **egymásra halmozzuk**.

```
# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Sex, fill=Smoke, y=after_stat(count/sum(count)))) +
  # oszlopdiagram kirajzolása
  geom_bar(col="#2e2e2e", position="stack") +
  # az x tengely beállítása
  scale_x_discrete(labels=c(Female="Nő", Male="Férfi")) +
  # a kitöltés (fill) paraméterek beállítása
  scale_fill_manual(values = c("#E9E9E9", "#C9C9C9", "#A9A9A9", "#898989"),
    na.translate = FALSE,
    labels=c(Never="Soha",
      Occas="Alkalmi",
      Regul="Rendszeres",
      Heavy="Erős")) +
```

```

# százalékos gyakoriságok megjelenítése
geom_text(stat = "count", size=2.6,
          position = position_stack(vjust = 0.5),
          aes(label =
              scales::percent(after_stat(count/sum(count)),
                              decimal.mark = ",",
                              accuracy = 0.1))) +

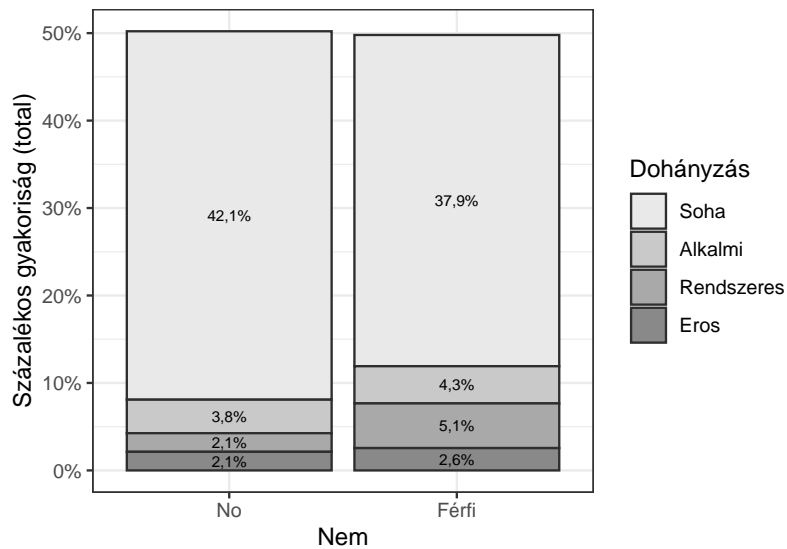
# y tengely formázása
scale_y_continuous(labels = scales::percent) +
# feliratok hozzáadása
labs(x = "Nem", y = "Százalékos gyakoriság (total)", fill="Dohányzás") +
# fekete-fehér stílus hozzáadása
theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/oszlopdiagram_04.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



5.2.10 Publikációkész ábra - 7.

Elkészítjük a publikációkész ábrát. Most olyan megjelenítést választottunk, amelyben **az egyes kategóriák százalékos relatív gyakoriságát jelenítjük meg a teljes mintára vonatkoztatva (a megjelenített százalékos értékek összege 100%)**. Az x tengelyen jelenítjük meg a dohányzási szokásokat, a kitöltés paraméterrel a nemeket különböztetjük meg, és az azonos dohányzási kategóriába tartozó százalékos gyakoriságokat **egymás mellett** jelenítjük meg.

```
# ggplot2 csomagok betöltése
library(ggplot2)

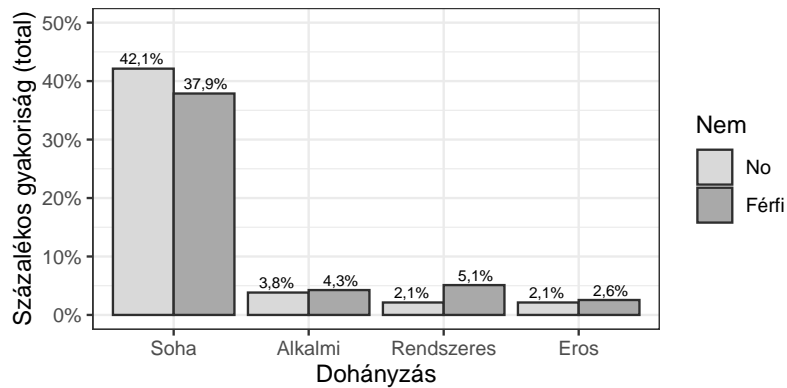
# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, fill=Sex, y=after_stat(count/sum(count)))) +
  # oszlopdiagram kirajzolása
  geom_bar(col="#2e2e2e", position="dodge") +
  # az x tengely beállítása
  scale_x_discrete(na.translate = FALSE,
                   labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +
  # a kitöltés (fill) paraméterek beállítása
  scale_fill_manual(values = c("#D9D9D9", "#A9A9A9"),
                   label=c(Female="Nő", Male="Férfi")) +
  # százalékos gyakoriságok megjelenítése
  geom_text(stat = "count", size=2.6,
           position=position_dodge(width=0.95), vjust=-0.4,
           aes(label =
               scales::percent(after_stat(count/sum(count)),
                               decimal.mark = ",",
                               accuracy = 0.1))) +

  # y tengely formázása
  scale_y_continuous(labels = scales::percent) +
  # y tengely láthatóságának beállítása
  coord_cartesian(ylim = c(0, 0.5)) +
  # feliratok hozzáadása
  labs(x = "Dohányzás", y = "Százalékos gyakoriság (total)", fill="Nem") +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása
```

```
# a p1 ábra mentése
ggsave("output/kep/oszlopdiagram_03.png", plot = p1,
       units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)
```

```
p1 # a p1 ábra kirajzolása a képernyőre
```



5.2.11 Publikációkész ábra - 8.

Elkészítjük a publikációkész ábrát. Most olyan megjelenítést választottunk, amelyben **az egyes kategóriák százalékos relatív gyakoriságát jelenítjük meg a teljes mintára vonatkoztatva** (a megjelenített százalékos értékek összege 100%). Az x tengelyen jelenítjük meg a nemeket, a kitöltés paraméterrel a dohányzási szokásokat különböztetjük meg, és az azonos nemhez tartozó százalékos gyakoriságokat **együtt** mellett jelenítjük meg.

```
# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Sex, fill=Smoke, y=after_stat(count/sum(count)))) +
  # oszlopdiagram kirajzolása
  geom_bar(col="#2e2e2e", position="dodge") +
  # az x tengely beállítása
  scale_x_discrete(labels=c(Female="Nő", Male="Férfi")) +
  # a kitöltés (fill) paraméterek beállítása
  scale_fill_manual(values = c("#E9E9E9", "#C9C9C9", "#A9A9A9", "#898989"),
                    na.translate = FALSE,
                    labels=c(Never="Soha",
```

```

        Occas="Alkalmi",
        Regul="Rendszeres",
        Heavy="Erős")) +
# százalékos gyakoriságok megjelenítése
geom_text(stat = "count", size=2.6,
          position=position_dodge(width=0.95), vjust=-0.4,
          aes(label =
              scales::percent(after_stat(count/sum(count)),
                              decimal.mark = ",",
                              accuracy = 0.1))) +

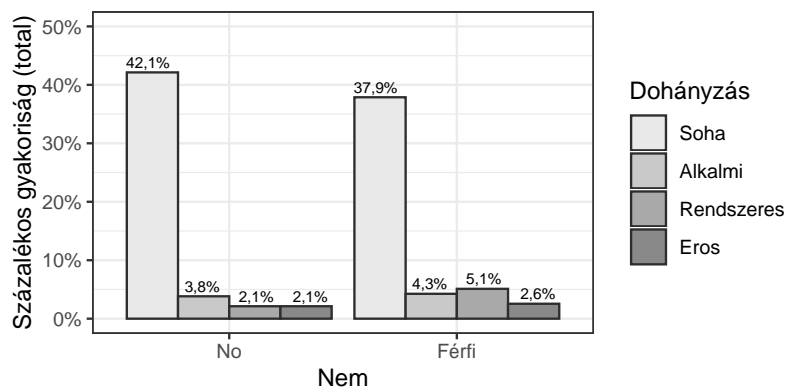
# y tengely formázása
scale_y_continuous(labels = scales::percent) +
# y tengely láthatóságának beállítása
coord_cartesian(ylim = c(0, 0.5)) +
# feliratok hozzáadása
labs(x = "Nem", y = "Százalékos gyakoriság (total)", fill="Dohányzás") +
# fekete-fehér stílus hozzáadása
theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/oszlopdiagram_04.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



5.2.12 Publikációkész ábra - 9.

Elkészítjük a publikációkész ábrát. Most olyan megjelenítést választottunk, amelyben **az egyes kategóriák százalékos relatív gyakoriságát jelenítjük meg az x tengelyen szereplő faktor egyes kategóriára vonatkoztatva (az x tengely minden egyes kategóriájában a megjelenített százalékos értékek összege 100%)**. Az x tengelyen jelenítjük meg a dohányzási szokásokat jelenítjük meg, a kitöltés paraméterrel a nemeket különböztetjük meg, és az azonos dohányzási szokáshoz tartozó százalékos gyakoriságokat **egymás mellett jelenítjük meg (ezek összege 100%)**.

```
# ggplot2 csomagok betöltése
library(ggplot2)
library(GGally)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, fill=Sex, by=Smoke, y=after_stat(prop))) +
  # oszlopdiaagram kirajzolása
  geom_bar(col="#2e2e2e", stat="prop", position="dodge") +
  # az x tengely beállítása
  scale_x_discrete(na.translate = FALSE,
                   labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +
  # a kitöltés (fill) paraméterek beállítása
  scale_fill_manual(values = c("#D9D9D9", "#898989"),
                   label=c(Female="Nő", Male="Férfi")) +
  # százalékos gyakoriságok megjelenítése
  geom_text(stat = "prop", size=2.6, col="black",
           position=position_dodge(width=0.95), vjust=-0.4,
           aes(label =
               scales::percent(after_stat(prop),
                               decimal.mark = ",",
                               accuracy = 0.1))) +

  # y tengely formázása
  scale_y_continuous(labels = scales::percent) +
  # y tengely láthatóságának beállítása
  coord_cartesian(ylim = c(0, 0.75)) +
  # feliratok hozzáadása
  labs(x = "Dohányzás", fill = "Nem",
       y = "Százalékos gyakoriság\n(dohányzási szokásonként vett)") +
  # fekete-fehér stílus hozzáadása
```

```

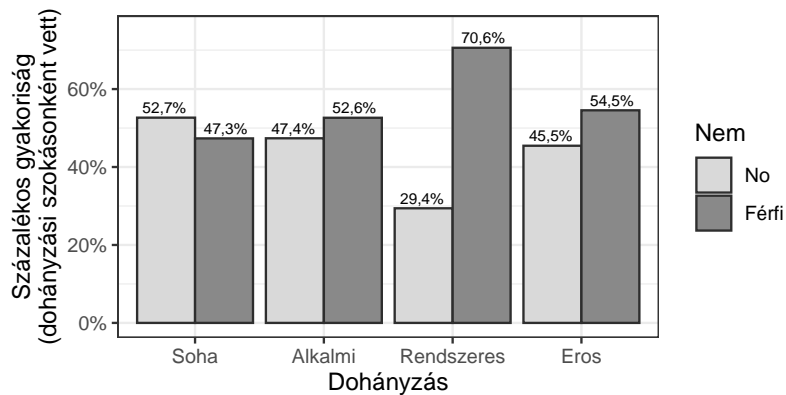
theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/oszlopdiagram_03.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



5.2.13 Publikációkész ábra - 10.

Elkészítjük a publikációkész ábrát. Most olyan megjelenítést választottunk, amelyben **az egyes kategóriák százalékos relatív gyakoriságát jelenítjük meg az x tengelyen szereplő faktor egyes kategóriára vonatkoztatva (az x tengely minden egyes kategóriájában a megjelenített százalékos értékek összege 100%)**. Az x tengelyen jelenítjük meg a nemeket, a kitöltés paraméterrel a dohányzási szokásokat különböztetjük meg, és az azonos nemhez tartozó százalékos gyakoriságokat **egymás mellett jelenítjük meg (ezek összege 100%)**.

```

# ggplot2 csomagok betöltése
library(ggplot2)
library(GGally)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Sex, fill=Smoke, by=Sex, y=after_stat(prop))) +
  # oszlopdiagram kirajzolása
  geom_bar(col="#2e2e2e", position="dodge", stat="prop") +

```

```

# az x tengely beállítása
scale_x_discrete(labels=c(Female="Nő", Male="Férfi")) +
# a kitöltés (fill) paraméterek beállítása
scale_fill_manual(values = c("#D9D9D9", "#B9B9B9", "#999999", "#797979"),
  na.translate = FALSE,
  labels=c(Never="Soha",
           Occas="Alkalmi",
           Regul="Rendszeres",
           Heavy="Erős")) +
# százalékos gyakoriságok megjelenítése
geom_text(stat = "prop", size=2.6, col="black",
  position=position_dodge(width=0.85), vjust=-0.4,
  aes(label =
    scales::percent(after_stat(prop),
                     decimal.mark = ",",
                     accuracy = 0.1))) +

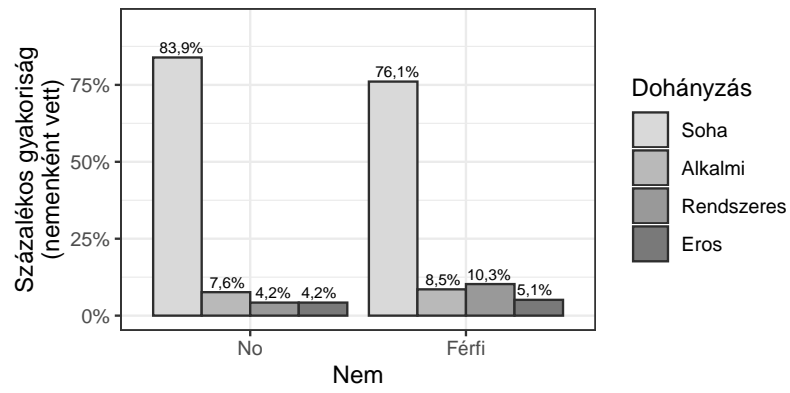
# y tengely formázása
scale_y_continuous(labels = scales::percent) +
# y tengely láthatóságának beállítása
coord_cartesian(ylim = c(0, 0.95)) +
# feliratok hozzáadása
labs(x = "Nem", fill="Dohányzás",
     y = "Százalékos gyakoriság\n(nemenként vett)") +
# fekete-fehér stílus hozzáadása
theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/oszlopdiagram_04.png", plot = p1,
  units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



6 Pontdiagram

6.1 Egydimenziós pontdiagram

i Feladat

A `survey.xlsx` 237 egyetemista adatát tartalmazza. A `Smoke` változó a dohányzásra vonatkozó kérdésre adott válaszokat tartalmazza. A `Pulse` változó a pulzust méri. Rajzoljuk meg a pulzus nyers pontjait a dohányzási szokás egyes kategóriáiban, adjuk meg az átlagot és a mediánt is.

6.1.1 Adatok beolvasása

Elvégezzük az adatok beolvasását a `survey.xlsx` fájlból. Ha az adatok SPSS `.sav` formátumban állnak rendelkezésre, akkor a beolvasás ugyanígy történik.

```
# Adatok beolvasása
d <- rio::import(file = "adat/survey.xlsx")
```

6.1.2 Előkészítés

Elképzeltető, hogy adataink közvetlenül a beolvasás után még nincsenek elemezhető állapotban. Néhány előkészítő lépést most valóban el kell végeznünk, de ez nem minden adatbázis esetén szükséges:

- a dohányzási szokásokat tartalmazó változó (`Smoke`) ordinális, így faktorrá kell alakítanunk
- rendezzük az ordinálissá alakított `Smoke` változó szintjeit a természetes sorrendbe.

```
library(tidyverse)
# faktorrá alakítás
d$Smoke <- factor(d$Smoke, ordered = T)
# faktor szintjeinek rendezése
```

```
d$Smoke <- forcats::fct_relevel(d$Smoke, "Never", "Occas", "Regul",
                                "Heavy")
```

6.1.3 Elemzés

Leíró statisztikai elemzés segítségével megnézzük, hogy milyen értékek ábrázolását fogjuk elvégezni.

```
# leíró statisztikai mutatók a jmv csomag segítségével
jmv::descriptives(data = d, vars = "Pulse", split = "Smoke")
#>
#> DESCRIPTIVES
#>
#> Descriptives
#>
#>
#>           Smoke      Pulse
#>
#> N
#>   Never      152
#>   Occas       16
#>   Regul       16
#>   Heavy        7
#> Missing
#>   Never       37
#>   Occas        3
#>   Regul        1
#>   Heavy        4
#> Mean
#>   Never  73.99342
#>   Occas  73.93750
#>   Regul  73.68750
#>   Heavy  78.28571
#> Median
#>   Never  72.00000
#>   Occas  76.00000
#>   Regul  69.00000
#>   Heavy  80.00000
#> Standard deviation
#>   Never  11.71035
#>   Occas   9.226547
#>   Regul  14.42553
#>   Heavy  11.70063
#> Minimum
#>   Never  35.00000
#>   Occas  60.00000
#>   Regul  50.00000
```

```

#>           Heavy      65.00000
#> Maximum      Never      104.0000
#>           Occas      87.00000
#>           Regul      104.0000
#>           Heavy      100.0000
#>

```

6.1.4 Publikációkész ábra - 1.

Elkészítjük a publikációkész ábrát. Most olyan megjelenítést választottunk, amelyben a pontokat egy kis eltolással jelenítjük meg, hogy az egymást átfedő pontok lehetőleg ne takarják ki egymást. A pontokat áttetszőre rajzoltuk, hogy az egymásra rajzolt pontok sűrűsége is látható legyen. Az átlagot és a mediánt is megjelenítjük. Az átlagot egy nagy szürke ponttal, a mediánt egy fekete vonallal jelöljük.

```

# ggplot2 csomagok betöltése
library(ggplot2)

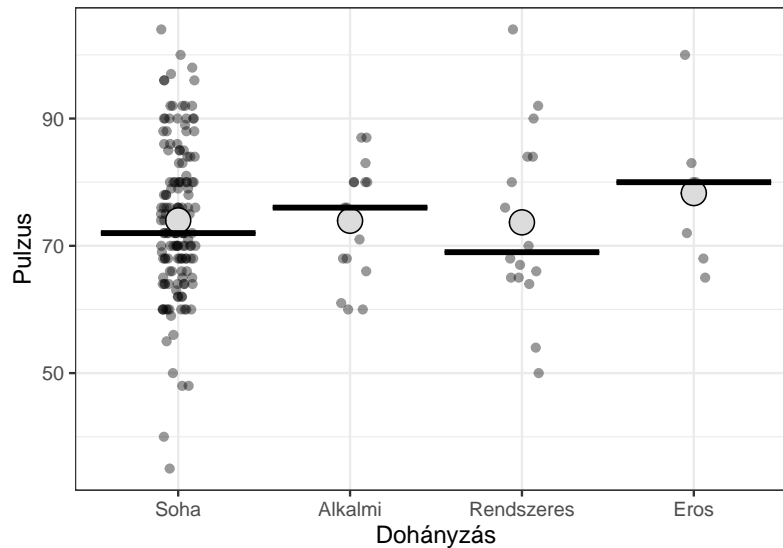
# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, y=Pulse)) +
  # pontdiagram kirajzolása, véletlen x irányú eltolással
  geom_jitter(width = 0.1, height = 0, alpha=0.4) +
  # átlag megjelenítése egy nagy szürke ponttal
  stat_summary(fun = mean, geom = "point",
               size=5, shape=21, fill="#dcdcdc", col="black",
               position = position_dodge(width = 0.5)) +
  # medián megjelenítése egy fekete vonallal
  stat_summary(fun=median, fun.min=median, fun.max=median,
               geom="crossbar", col="black",
               position = position_dodge(width = 0.5)) +
  # x tengely beállítása
  scale_x_discrete(na.translate = FALSE,
                  labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +
  # feliratok hozzáadása
  labs(x = "Dohányzás", y = "Pulzus") +
  # fekete-fehér stílus hozzáadása
  theme_bw()

```

```
# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/pontdiagram_01.png", plot = p1,
       units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre
```



6.1.5 Publikációkész ábra - 2.

Elkészítjük a publikációkész ábrát. Most olyan megjelenítést választottunk, amely a pontok egymásra rajzolását a lehetőség szerint elkerüli. A pontokat áttetszőre rajzoltuk, hogy az egymásra rajzolt pontok sűrűsége is látható legyen. Az átlagot és a mediánt is megjelenítjük. Az átlagot egy nagy szürke ponttal, a mediánt egy fekete vonallal jelöljük.

```
# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, y=Pulse)) +
  # pontok kirajzolása, egymásra rajzolás elkerülése
  ggbeeswarm::geom_quasirandom(alpha=0.4) +
  # átlag megjelenítése egy nagy szürke ponttal
```



```

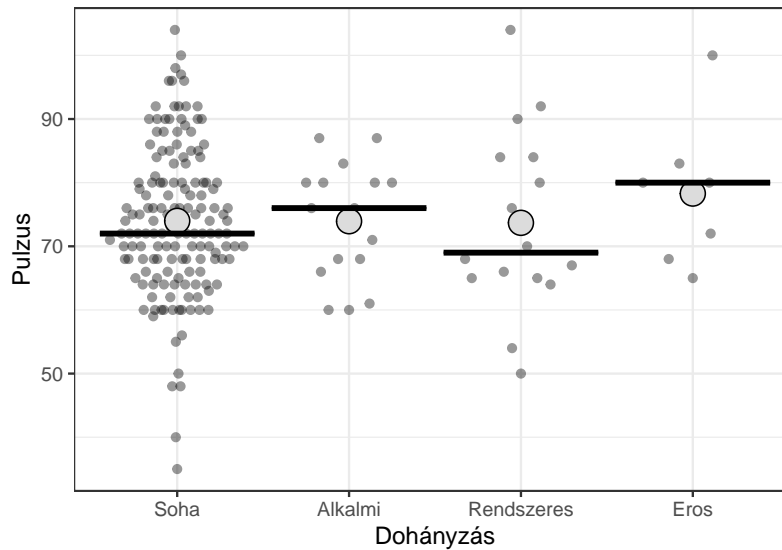
stat_summary(fun = mean, geom = "point",
              size=5, shape=21, fill="#dcdcdc", col="black",
              position = position_dodge(width = 0.5)) +
# medián megjelenítése egy fekete vonallal
stat_summary(fun=median, fun.min=median, fun.max=median,
              geom="crossbar", col="black",
              position = position_dodge(width = 0.5)) +
# x tengely beállítása
scale_x_discrete(na.translate = FALSE,
                  labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +
# feliratok hozzáadása
labs(x = "Dohányzás", y = "Pulzus") +
# fekete-fehér stílus hozzáadása
theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/pontdiagram_02.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



6.2 Kétdimenziós pontdiagram - csoportosítás nélkül

i Feladat

A `survey.xlsx` 237 egyetemista adatát tartalmazza. A `Height` változó a magasságot tartalmazza cm-ben, míg a `NW.Hnd` annak a kezünknek a méretét, amelyikkel nem szoktunk írni. Készítsünk kétdimenziós pontdiagramot a testmagasság és a kéz hossza változók kapcsolatának vizsgálatára.

6.2.1 Adatok beolvasása

Elvégezzük az adatok beolvasását a `survey.xlsx` fájlból. Ha az adatok SPSS `.sav` formátumban állnak rendelkezésre, akkor a beolvasás ugyanígy történik.

```
# Adatok beolvasása
d <- rio::import(file = "adat/survey.xlsx")
```

6.2.2 Előkészítés

Elképzeltető, hogy adataink közvetlenül a beolvasás után még nincsenek elemezhető állapotban. Azonban ez nem minden adatbázis esetén szükséges, ebben az esetben sem.

6.2.3 Elemzés

Leíró statisztikai elemzés segítségével megnézzük, hogy milyen értékek ábrázolását fogjuk elvégezni.

```
# leíró statisztikai mutatók a jmv csomag segítségével
jmv::descriptives(data = d, vars = c("Height", "NW.Hnd"))
#>
#> DESCRIPTIVES
#>
#> Descriptives
#>
#>
#>           Height      NW.Hnd
#>
#> N                209        236
#> Missing           28         1
#> Mean            172.3809    18.58263
#> Median           171.0000    18.50000
#> Standard deviation 9.847528    1.967068
#> Minimum           150.0000    12.50000
#> Maximum           200.0000    23.50000
#>
```

6.2.4 Publikációkész ábra - 1.

Elkészítjük a publikációkész ábrát. Most csupán a pontokat jelenítjük meg.

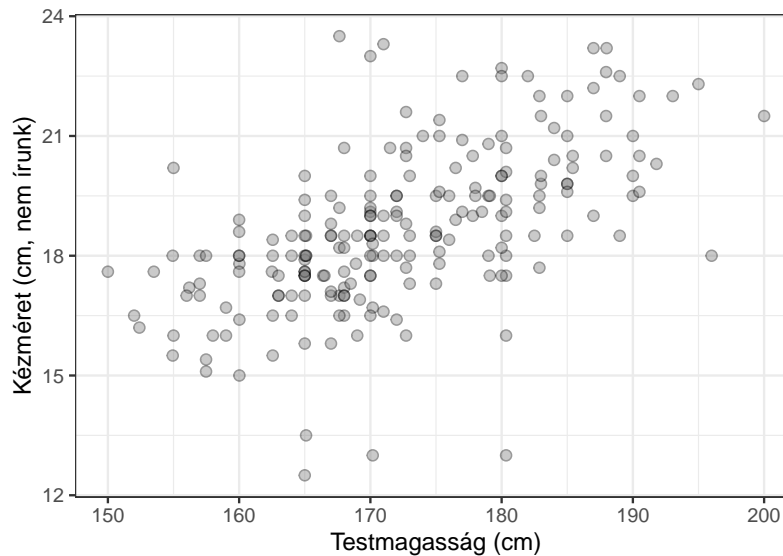
```
# ggplot2 csomagok betöltése
library(ggplot2)

# p1 - kétdimenziós pontdiagram
p1 <- ggplot(data=d, aes(x=Height, y=NW.Hnd)) +
  # pontok kirajzolása
  geom_point(shape=21, size=2, fill="#7c7c7c", col="black", alpha=0.4) +
  # feliratok hozzáadása
  labs(x = "Testmagasság (cm)", y = "Kézméret (cm, nem írunk)") +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása
```

```
# a p1 ábra mentése
ggsave("output/kep/pontdiagram_03.png", plot = p1,
       units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)
```

```
p1 # a p1 ábra kirajzolása a képernyőre
```



6.2.5 Publikációkész ábra - 2.

Elkészítjük a publikációkész ábrát. Most a pontok mellett a lineáris regressziós egyenest is megjelenítjük.

```
# ggplot2 csomagok betöltése
library(ggplot2)

# p1 - kétdimenziós pontdiagram
p1 <- ggplot(data=d, aes(x=Height, y=NW.Hnd)) +
  # pontok kirajzolása
  geom_point(shape=21, size=2, fill="#7c7c7c", col="black", alpha=0.4) +
  # a regressziós egyenes
  geom_smooth(method = "lm", se=F, colour="black", linewidth=0.8) +
  # feliratok hozzáadása
  labs(x = "Testmagasság (cm)", y = "Kézméret (cm, nem írunk)") +
  # fekete-fehér stílus hozzáadása
```

```

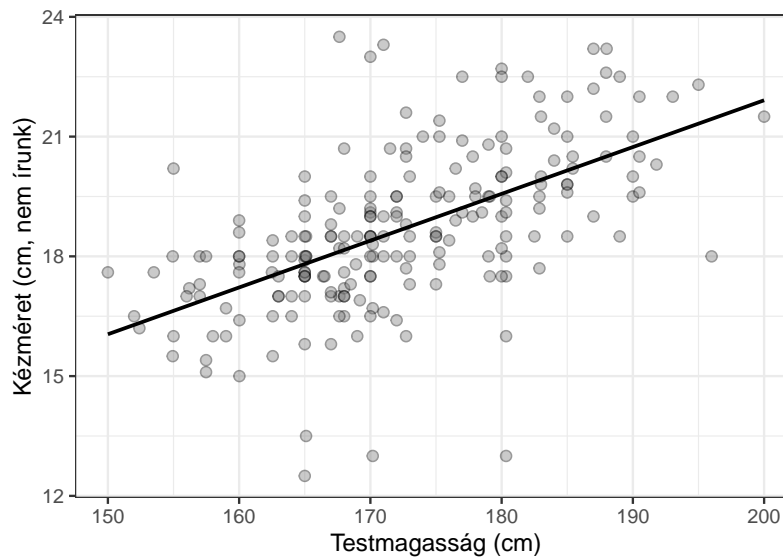
theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/pontdiagram_04.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



6.2.6 Publikációkész ábra - 3.

Elkészítjük a publikációkész ábrát. Most pontok és a lineáris regressziós egyenes megjelenítése mellett egy görbeillesztést is elvégezzünk.

```

# ggplot2 csomagok betöltése
library(ggplot2)

# p1 - kétdimenziós pontdiagram
p1 <- ggplot(data=d, aes(x=Height, y=NW.Hnd)) +
  # pontok kirajzolása
  geom_point(shape=21, size=2, fill="#7c7c7c", col="black", alpha=0.4) +
  # a regressziós egyenes

```

```

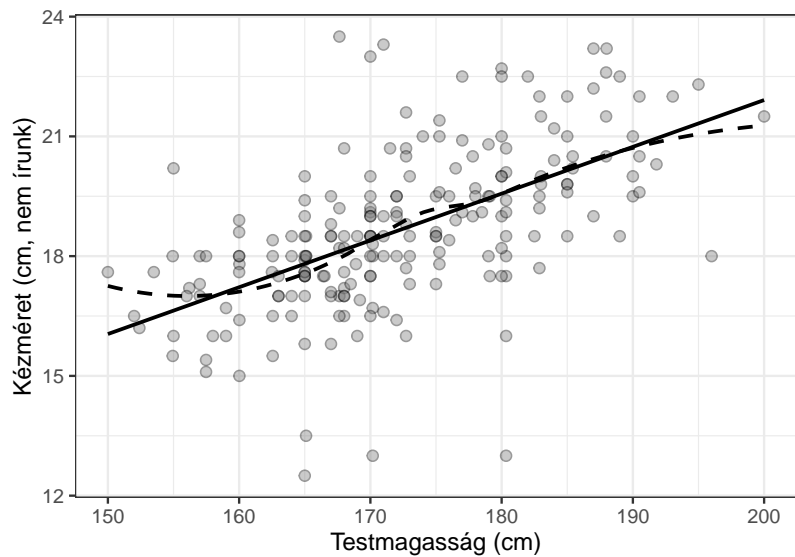
geom_smooth(method = "lm", se=F, colour="black", linewidth=0.8) +
# illesztés görbe segítségével
geom_smooth(method = "loess", se=F, colour="black", linetype="dashed", size=0.8) +
# feliratok hozzáadása
labs(x = "Testmagasság (cm)", y = "Kézméret (cm, nem írunk)") +
# fekete-fehér stílus hozzáadása
theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/pontdiagram_05.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



6.3 Kétdimenziós pontdiagram - csoportosítással

i Feladat

A `survey.xlsx` 237 egyetemista adatát tartalmazza. A `Height` változó a magasságot tartalmazza cm-ben, míg a `NW.Hnd` annak a kezünknek a méretét, amelyekkel nem szoktunk írni. A nem változó (`Sex`) figyelembevételével készítsünk kétdimenziós pontdiagramot a testmagasság és a kéz hossza változók kapcsolatának vizsgálatára.

6.3.1 Adatok beolvasása

Elvégezzük az adatok beolvasását a `survey.xlsx` fájlból. Ha az adatok SPSS `.sav` formátumban állnak rendelkezésre, akkor a beolvasás ugyanígy történik.

```
# Adatok beolvasása
d <- rio::import(file = "adat/survey.xlsx")
```

6.3.2 Előkészítés

Elképzelhető, hogy adataink közvetlenül a beolvasás után még nincsenek elemezhető állapotban. Néhány előkészítő lépést most valóban el kell végeznünk, de ez nem minden adatbázis esetén szükséges:

- a nem változó (`Sex`) nominális, így faktorrá kell alakítanunk,
- azokkal a személyekkel nem szeretnénk elemzést végezni, akiknek nincs megadva a nemük, így ezeket a sorokat kiszűrjük az adatbázisból.

```
library(tidyverse)
# faktorrá alakítás
d$Sex <- factor(d$Sex)
# kiszűrjük azokat a személyeket, akiknek nincs megadva a
# nemük
d <- d |>
  dplyr::filter(!is.na(Sex))
```

6.3.3 Elemzés

Leíró statisztikai elemzés segítségével megnézzük, hogy milyen értékek ábrázolását fogjuk elvégezni.

```
# leíró statisztikai mutatók a jmv csomag segítségével
jmv::descriptives(data = d, vars = "Sex", freq = T)
```

```
#>
#> DESCRIPTIVES
#>
#> Descriptives
#>
#>                               Sex
#>
#>      N                236
#>      Missing            0
#>      Mean
#>      Median
#>      Standard deviation
#>      Minimum
#>      Maximum
#>
#>
#>
#> FREQUENCIES
#>
#> Frequencies of Sex
#>
#>      Sex      Counts    % of Total    Cumulative %
#>
#>      Female      118      50.00000      50.00000
#>      Male        118      50.00000      100.00000
#>
```

```
# leíró statisztikai mutatók a jmv csomag segítségével
jmv::descriptives(data = d, vars = c("Height", "NW.Hnd"), splitBy = "Sex")
```

```
#>
#> DESCRIPTIVES
#>
#> Descriptives ...
#>
#>                               ...
#>                               Sex      Height    NW.H...
#>                               ...
#>      N                Female      102      ...
#>                Male      106      ...
#>      Missing      Female      16      ...
```



```

#>           Male           12           ...
#>   Mean           Female    165.6867    17.4...
#>           Male           178.8260    19.7...
#>   Median          Female    166.7500    17.6...
#>           Male           180.0000    19.5...
#>   Standard deviation Female    6.151777    1.40...
#>           Male           8.380252    1.80...
#>   Minimum          Female    150.0000    12.5...
#>           Male           154.9400    13.3...
#>   Maximum          Female    180.3400    20.7...
#>           Male           200.0000    23.5...
#>           ...

```

6.3.4 Publikációkész ábra - 1.

Elkészítjük a publikációkész ábrát. Most csupán a pontokat jelenítjük meg, de a nem változó hatását is figyelembe vesszük.

```

# ggplot2 csomagok betöltése
library(ggplot2)

# p1 - kétdimenziós pontdiagram
p1 <- ggplot(data=d, aes(x=Height, y=NW.Hnd,
                        col=Sex, shape=Sex)) +
  # pontok kirajzolása
  geom_point(size=2) +
  # színek beállítása
  scale_colour_manual(values=c("#acacac", "#3d3d3d"),
                      labels=c(Female="Nő", Male="Férfi")) +
  # formák beállítása
  scale_shape_discrete(labels=c(Female="Nő", Male="Férfi")) +
  # feliratok hozzáadása
  labs(x = "Testmagasság (cm)", y = "Kézméret (cm, nem írunk)",
       color="Nem", shape="Nem") +
  # fekete-fehér stílus hozzáadása
  theme_bw()

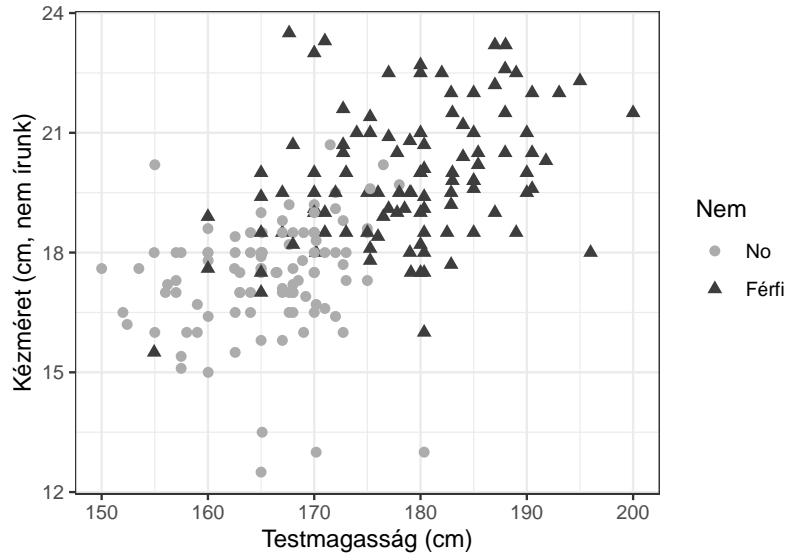
# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/pontdiagram_06.png", plot = p1,

```

```
units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)
```

```
p1 # a p1 ábra kirajzolása a képernyőre
```



6.3.5 Publikációkész ábra - 2.

Elkészítjük a publikációkész ábrát. Most pontok mellett a lineáris regressziós egyenest is megjelenítjük. A nem változó hatását is figyelembe vesszük.

```
# ggplot2 csomagok betöltése
library(ggplot2)

# p1 - kétdimenziós pontdiagram
p1 <- ggplot(data=d, aes(x=Height, y=NW.Hnd,
                        col=Sex, shape=Sex, linetype=Sex)) +
  # pontok kirajzolása
  geom_point(size=2) +
  # színek beállítása
  scale_colour_manual(values=c("#acacac", "#3d3d3d"),
    labels=c(Female="Nő", Male="Férfi")) +
  # formák beállítása
  scale_shape_discrete(labels=c(Female="Nő", Male="Férfi")) +
```

```

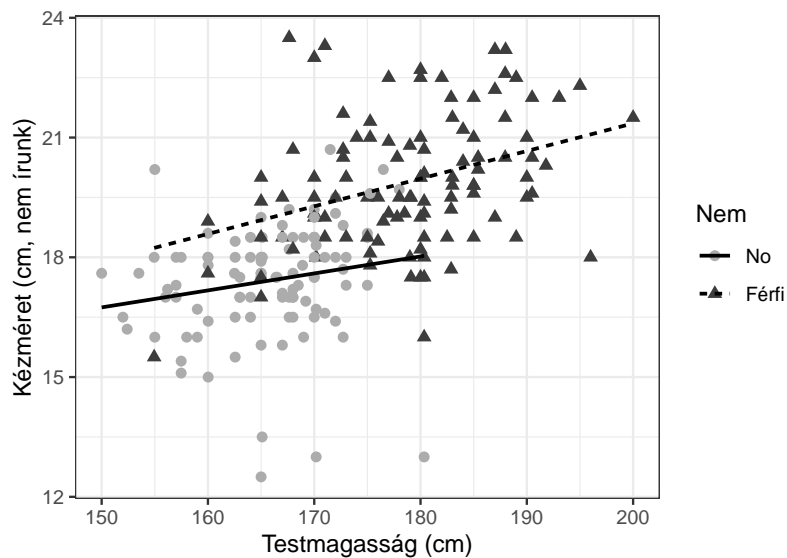
# vonaltípus beállítása
scale_linetype_discrete(labels=c(Female="Nő", Male="Férfi")) +
# a regressziós egyenes
geom_smooth(method = "lm", se=F, colour="black", linewidth=0.8) +
# feliratok hozzáadása
labs(x = "Testmagasság (cm)", y = "Kézméret (cm, nem írunk)",
     color="Nem", shape="Nem", linetype="Nem") +
# fekete-fehér stílus hozzáadása
theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/pontdiagram_07.png", plot = p1,
       units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



6.3.6 Publikációkész ábra - 3.

Elkészítjük a publikációkész ábrát. Most pontok és a lineáris regressziós egyenes megjelenítése mellett egy görbeillesztést is elvégzünk. A nem változó hatását is figyelembe vesszük.

```

# ggplot2 csomagok betöltése
library(ggplot2)

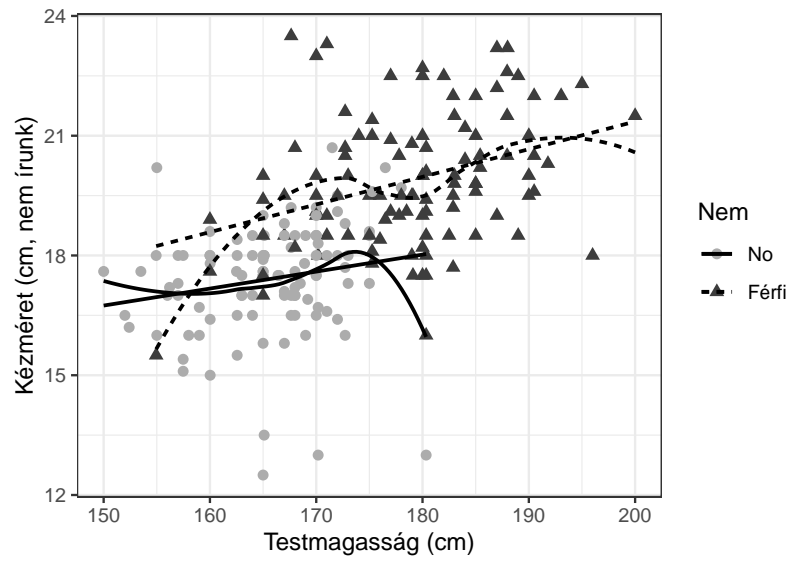
# p1 - kétdimenziós pontdiagram
p1 <- ggplot(data=d, aes(x=Height, y=NW.Hnd,
                        col=Sex, shape=Sex, linetype=Sex)) +
  # pontok kirajzolása
  geom_point(size=2) +
  # színek beállítása
  scale_colour_manual(values=c("#acacac", "#3d3d3d"),
    labels=c(Female="Nő", Male="Férfi")) +
  # formák beállítása
  scale_shape_discrete(labels=c(Female="Nő", Male="Férfi")) +
  # vonaltípus beállítása
  scale_linetype_discrete(labels=c(Female="Nő", Male="Férfi")) +
  # a regressziós egyenes
  geom_smooth(method = "lm", se=F, colour="black", linewidth=0.8) +
  # illesztés görbe segítségével
  geom_smooth(method = "loess", se=F, colour="black", size=0.8) +
  # feliratok hozzáadása
  labs(x = "Testmagasság (cm)", y = "Kézméret (cm, nem írunk)",
    color="Nem", shape="Nem", linetype="Nem") +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/pontdiagram_08.png", plot = p1,
  units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



7 Átlagábrák

7.1 Átlagok egy faktor esetén

Feladat

A `survey.xlsx` 237 egyetemista adatát tartalmazza. A `Smoke` változó a dohányzásra vonatkozó kérdésre adott válaszokat tartalmazza. A `Pulse` változó a pulzust méri. Adjuk meg a pulzus átlagát a dohányzási szokás egyes csoportjaiban. Jelenítsük meg a 95%-os konfidencia intervallumot is.

7.1.1 Adatok beolvasása

Elvégezzük az adatok beolvasását a `survey.xlsx` fájlból. Ha az adatok SPSS `.sav` formátumban állnak rendelkezésre, akkor a beolvasás ugyanígy történik.

```
# Adatok beolvasása
d <- rio::import(file = "adat/survey.xlsx")
```

7.1.2 Előkészítés

Elképzelhető, hogy adataink közvetlenül a beolvasás után még nincsenek elemezhető állapotban. Néhány előkészítő lépést most valóban el kell végeznünk, de ez nem minden adatbázis esetén szükséges:

- a dohányzási szokásokat tartalmazó változó (`Smoke`) ordinális, így faktorrá kell alakítanunk
- rendezzük az ordinálissá alakított `Smoke` változó szintjeit a természetes sorrendbe.

```
library(tidyverse)
# faktorrá alakítás
d$Smoke <- factor(d$Smoke, ordered = T)
# faktor szintjeinek rendezése
```

```
d$Smoke <- forcats::fct_relevel(d$Smoke, "Never", "Occas", "Regul",
                                "Heavy")
```

7.1.3 Elemzés

Leíró statisztikai elemzés segítségével megnézzük, hogy milyen értékek ábrázolását fogjuk elvégezni.

```
# leíró statisztikai mutatók a jmv csomag segítségével
jmv::descriptives(data = d, vars = "Pulse", split = "Smoke")
#>
#> DESCRIPTIVES
#>
#> Descriptives
#>
#>           Smoke      Pulse
#>
#>      N           Never      152
#>           Occas         16
#>           Regul         16
#>           Heavy          7
#>      Missing      Never      37
#>           Occas          3
#>           Regul          1
#>           Heavy          4
#>      Mean           Never  73.99342
#>           Occas      73.93750
#>           Regul      73.68750
#>           Heavy      78.28571
#>      Median          Never  72.00000
#>           Occas      76.00000
#>           Regul      69.00000
#>           Heavy      80.00000
#>      Standard deviation      Never  11.71035
#>           Occas      9.226547
#>           Regul      14.42553
#>           Heavy      11.70063
#>      Minimum          Never  35.00000
#>           Occas      60.00000
#>           Regul      50.00000
```

```

#>           Heavy      65.00000
#> Maximum      Never      104.0000
#>           Occas      87.00000
#>           Regul      104.0000
#>           Heavy      100.0000
#>

```

7.1.4 Publikációkész ábra - 1.

Elkészítjük a publikációkész ábrát. Most oszlopdiaqramon ábrázoljuk az átlagokat és a hozzájuk tartozó 95%-os konfidencia intervallumot. Ehhez a `pointrange` geometriai elemet hívjuk segítségül.

```

# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, y=Pulse)) +
  # oszlop kirajzolása
  stat_summary(fun=mean, geom="bar", fill="#dcdcdc", colour="black") +
  # hozzá tartozó konfidencia intervallum kirajzolása, pointrange
  stat_summary(fun.data=mean_cl_normal, geom="pointrange") +
  # az y tengely látható tartományának beállítása
  coord_cartesian(ylim = c(0, 100)) +
  # az x tengely beállítása
  scale_x_discrete(na.translate = FALSE,
                   labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +

  # feliratok hozzáadása
  labs(x = "Dohányzás", y = "Pulzus") +
  # fekete-fehér stílus hozzáadása
  theme_bw()

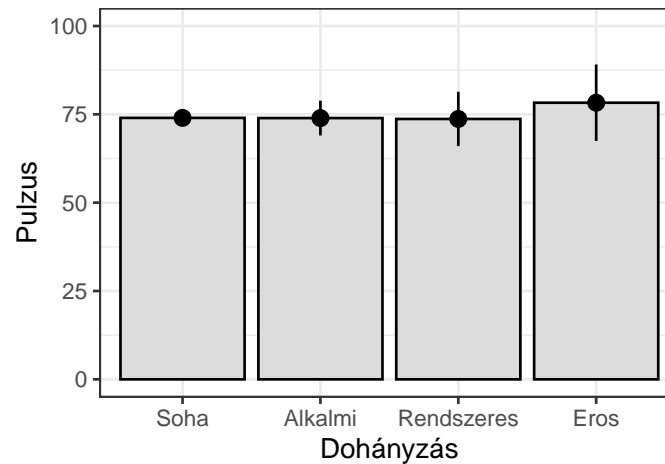
# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/atlagabra_01.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

```



```
p1 # a p1 ábra kirajzolása a képernyőre
```



7.1.5 Publikációkész ábra - 2.

Elkészítjük a publikációkész ábrát. Most oszlopdiaagramon ábrázoljuk az átlagokat és a hozzájuk tartozó 95%-os konfidencia intervallumot. Ehhez az `errorbar` geometriai elemet hívjuk segítségül.

```
# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, y=Pulse)) +
  # oszlop kirajzolása
  stat_summary(fun=mean, geom="bar", fill="#dcdcdc", colour="black") +
  # hozzá tartozó konfidencia intervallum kirajzolása, errorbar
  stat_summary(fun.data=mean_cl_normal, geom="errorbar", width=0.2) +
  # az y tengely látható tartományának beállítása
  coord_cartesian(ylim = c(0, 100)) +
  # az x tengely beállítása
  scale_x_discrete(na.translate = FALSE,
                   labels=c(Never="Soha",
                             Occas="Alkalmi",
                             Regul="Rendszeres",
                             Heavy="Erős")) +
  # feliratok hozzáadása
```

```

labs(x = "Dohányzás", y = "Pulzus") +
# fekete-fehér stílus hozzáadása
theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/atlagabra_02.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



7.1.6 Publikációkész ábra - 3.

Elkészítjük a publikációkész ábrát. Most pontokkal ábrázoljuk az átlagokat. A 95%-os konfidencia intervallumot az `errorbar` segítségével hozzuk létre.

```

# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, y=Pulse)) +
  # konfidencia intervallum kirajzolása, errorbar
  stat_summary(fun.data=mean_cl_normal, geom="errorbar", width=0.2) +
  # átlagok kirajzolása, pontokkal

```

```

stat_summary(fun=mean, geom="point",
              size=2, shape=21, fill="#dcdcdc", colour="black") +
# az y tengely látható tartományának beállítása
coord_cartesian(ylim = c(0, 100)) +
# az x tengely beállítása
scale_x_discrete(na.translate = FALSE,
                  labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +

# feliratok hozzáadása
labs(x = "Dohányzás", y = "Pulzus") +
# fekete-fehér stílus hozzáadása
theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/atlagabra_03.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



7.1.7 Publikációkész ábra - 4.

Elkészítjük a publikációkész ábrát. Most pontokkal ábrázoljuk az átlagokat. A 95%-os konfidencia intervallumot az `errorbar` segítségével hozzuk létre. Az átlagokat vonallal is összekötjük.

```
# ggplot2 csomagok betöltése
library(ggplot2)

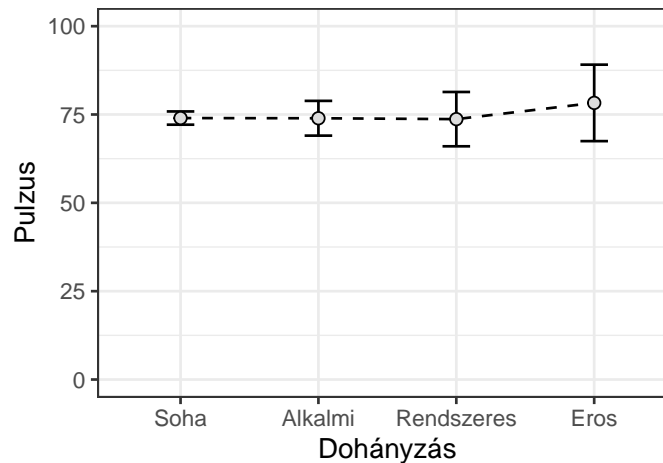
# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, y=Pulse)) +
  # konfidencia intervallum kirajzolása, errorbar
  stat_summary(fun.data=mean_cl_normal, geom="errorbar", width=0.2) +
  # vonal megrajzolása
  stat_summary(fun=mean, geom="line", aes(group=1), colour="black", linetype="dashed") +
  # átlagok kirajzolása, pontokkal
  stat_summary(fun=mean, geom="point",
               size=2, shape=21, fill="#dcdcdc", colour="black") +
  # az y tengely látható tartományának beállítása
  coord_cartesian(ylim = c(0, 100)) +
  # az x tengely beállítása
  scale_x_discrete(na.translate = FALSE,
                  labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +

  # feliratok hozzáadása
  labs(x = "Dohányzás", y = "Pulzus") +
  # fekete-fehér stílus hozzáadása
  theme_bw()

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/atlagabra_04.png", plot = p1,
       units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre
```



7.2 Átlagok két faktor esetén

i Feladat

A `survey.xlsx` 237 egyetemista adatát tartalmazza. A `Smoke` változó a dohányzásra vonatkozó kérdésre adott válaszokat tartalmazza, a `Sex` változó a nemre vonatkozó információt tárolja. A `Pulse` változó a pulzust méri. Adjuk meg a pulzus átlagát a dohányzási szokás és a nem egyes csoportjaiban. Jelenítsük meg a 95%-os konfidencia intervallumot is.

7.2.1 Adatok beolvasása

Elvégezzük az adatok beolvasását a `survey.xlsx` fájlból. Ha az adatok SPSS `.sav` formátumban állnak rendelkezésre, akkor a beolvasás ugyanígy történik.

```
# Adatok beolvasása
d <- rio::import(file = "adat/survey.xlsx")
```

7.2.2 Előkészítés

Elképzeltető, hogy adataink közvetlenül a beolvasás után még nincsenek elemezhető állapotban. Néhány előkészítő lépést most valóban el kell végeznünk, de ez nem minden adatbázis esetén szükséges:

- a nem változó (`Sex`) nominális, így faktorrá kell alakítanunk,

- a dohányzási szokásokat tartalmazó változó (**Smoke**) ordinális, így faktorrá kell alakítanunk,
- rendezzük az ordinálissá alakított **Smoke** változó szintjeit a természetes sorrendbe,
- azokkal a személyekkel nem szeretnénk elemzést végezni, akiknek nincs megadva a nemük, így ezeket a sorokat kiszűrjük az adatbázisból.

```
library(tidyverse)
# faktorrá alakítás
d$Sex <- factor(d$Sex)
# faktorrá alakítás
d$Smoke <- factor(d$Smoke)
# faktor szintjeinek rendezése
d$Smoke <- forcats::fct_relevel(d$Smoke, "Never", "Occas", "Regul",
                                "Heavy")
# kiszűrjük azokat a személyeket, akiknek nincs megadva a
# nemük
d <- d |>
  dplyr::filter(!is.na(Sex))
```

7.2.3 Elemzés

Leíró statisztikai elemzés segítségével megnézzük, hogy milyen értékek ábrázolását fogjuk elvégezni.

```
# leíró statisztikai mutatók a jmv csomag segítségével
jmv::descriptives(data = d, vars = "Pulse", split = c("Smoke",
  "Sex"))

#>
#> DESCRIPTIVES
#>
#> Descriptives ...
#>
#> ...
#>      Smoke      Sex      Pulse ...
#>      ...
#>      N      Never      Female      7...
#>      Male      7...
#>      Occas      Female      ...
#>      Male      ...
#>      Regul      Female      ...
#>      Male      1...
#>      Heavy      Female      ...
```

#>			Male	...
#>	Missing	Never	Female	2...
#>			Male	1...
#>		Occas	Female	...
#>			Male	...
#>		Regul	Female	...
#>			Male	...
#>		Heavy	Female	...
#>			Male	...
#>	Mean	Never	Female	75.6075...
#>			Male	72.2361...
#>		Occas	Female	73.3750...
#>			Male	74.5000...
#>		Regul	Female	69.2500...
#>			Male	75.1666...
#>		Heavy	Female	75.0000...
#>			Male	82.6666...
#>	Median	Never	Female	75.0000...
#>			Male	72.0000...
#>		Occas	Female	78.0000...
#>			Male	73.5000...
#>		Regul	Female	67.5000...
#>			Male	72.0000...
#>		Heavy	Female	76.0000...
#>			Male	83.0000...
#>	Standard deviation	Never	Female	11.4078...
#>			Male	11.9429...
#>		Occas	Female	11.2369...
#>			Male	7.44503...
#>		Regul	Female	17.3853...
#>			Male	13.8487...
#>		Heavy	Female	6.00000...
#>			Male	17.5023...
#>	Minimum	Never	Female	40.0000...
#>			Male	35.0000...
#>		Occas	Female	60.0000...
#>			Male	66.0000...
#>		Regul	Female	50.0000...
#>			Male	54.0000...
#>		Heavy	Female	68.0000...
#>			Male	65.0000...
#>	Maximum	Never	Female	104.000...

```

#>
#>           Occas      Male      97.0000...
#>           Female      87.0000...
#>           Male      87.0000...
#>           Regul      Female      92.0000...
#>           Male      104.000...
#>           Heavy      Female      80.0000...
#>           Male      100.000...
#>

```

7.2.4 Publikációkész ábra - 1.

Elkészítjük a publikációkész ábrát. Most oszlopdiaagramon ábrázoljuk az átlagokat és a hozzájuk tartozó 95%-os konfidencia intervallumot. Ehhez a `pointrange` geometriai elemet hívjuk segítségül.

```

# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, y=Pulse, fill=Sex)) +
  # oszlop kirajzolása
  stat_summary(fun=mean, geom="bar", position="dodge", col="#656565") +
  # hozzá tartozó konfidencia intervallum kirajzolása, pointrange
  stat_summary(fun.data=mean_cl_normal, geom="pointrange",
    position=position_dodge(width=0.90), show_guide = FALSE) +
  # az y tengely látható tartományának beállítása
  coord_cartesian(ylim = c(0, 100)) +
  # az x tengely beállítása
  scale_x_discrete(na.translate = FALSE,
    labels=c(Never="Soha",
      Occas="Alkalmi",
      Regul="Rendszeres",
      Heavy="Erős")) +
  # a kitöltés (fill) paraméterek beállítása
  scale_fill_manual(values = c("#DCDCDC", "#A9A9A9"),
    label=c(Female="Nő", Male="Férfi")) +
  # feliratok hozzáadása
  labs(x = "Dohányzás", y = "Pulzus", fill="Nem") +
  # fekete-fehér stílus hozzáadása
  theme_bw() +

```



```

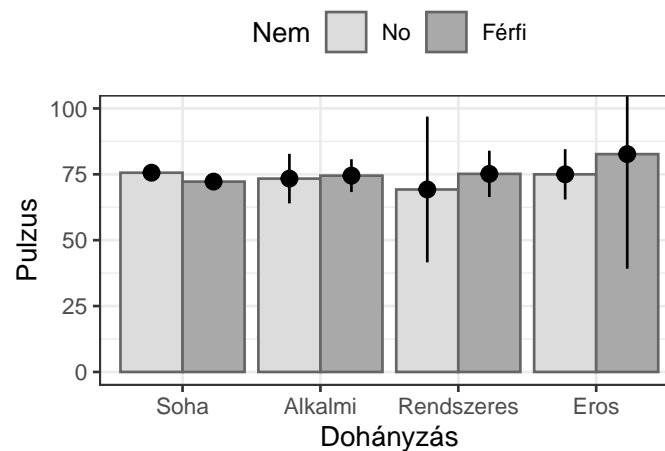
# a jelmagyarázat pozíciójának beállítása
theme(legend.position = "top")

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/atlagabra_05.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



7.2.5 Publikációkész ábra - 2.

Elkészítjük a publikációkész ábrát. Most oszlopdiagramon ábrázoljuk az átlagokat és a hozzájuk tartozó 95%-os konfidencia intervallumot. Ehhez az `errorbar` geometriai elemet hívjuk segítségül.

```

# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, y=Pulse, fill=Sex)) +
  # oszlop kirajzolása
  stat_summary(fun=mean, geom="bar", position="dodge", col="#656565") +
  # hozzá tartozó konfidencia intervallum kirajzolása, errorbar

```

```

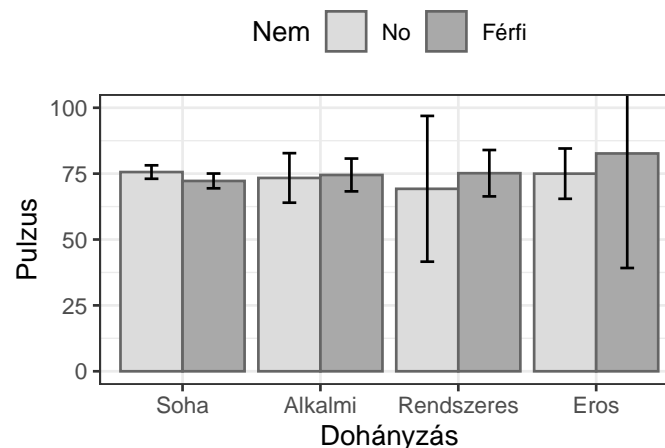
stat_summary(fun.data=mean_cl_normal, geom="errorbar", width=0.2,
              position=position_dodge(width=0.90), show_guide = FALSE) +
# az y tengely látható tartományának beállítása
coord_cartesian(ylim = c(0, 100)) +
# az x tengely beállítása
scale_x_discrete(na.translate = FALSE,
                  labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +
# a kitöltés (fill) paraméterek beállítása
scale_fill_manual(values = c("#DCDCDC", "#A9A9A9"),
                  label=c(Female="Nő", Male="Férfi")) +
# feliratok hozzáadása
labs(x = "Dohányzás", y = "Pulzus", fill="Nem") +
# fekete-fehér stílus hozzáadása
theme_bw() +
# a jelmagyarázat pozíciójának beállítása
theme(legend.position = "top")

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/atlagabra_06.png", plot = p1,
        units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```



7.2.6 Publikációkész ábra - 3.

Elkészítjük a publikációkész ábrát. Most pontokkal ábrázoljuk az átlagokat. A 95%-os konfidencia intervallumot az `errorbar` segítségével hozzuk létre.

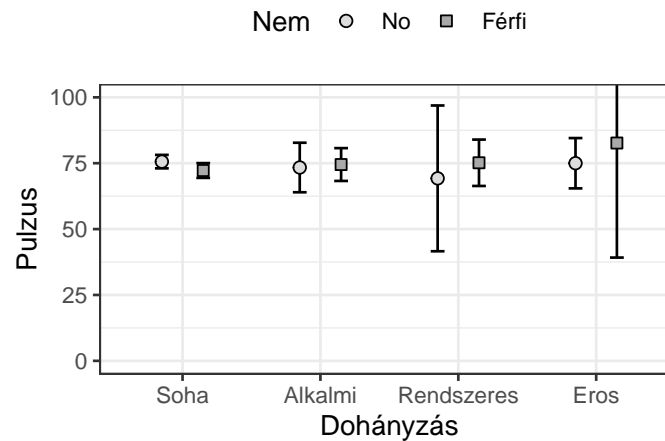
```
# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, y=Pulse, fill=Sex, shape=Sex)) +
  # hozzá tartozó konfidencia intervallum kirajzolása, errorbar
  stat_summary(fun.data=mean_cl_normal, geom="errorbar", width=0.2,
               position=position_dodge(width=0.6), show_guide = FALSE) +
  # átlagok kirajzolása, pontokkal
  stat_summary(fun=mean, geom="point", position=position_dodge(width=0.6),
               size=2) +
  # az y tengely látható tartományának beállítása
  coord_cartesian(ylim = c(0, 100)) +
  # az x tengely beállítása
  scale_x_discrete(na.translate = FALSE,
                  labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +
  # a kitöltés (fill) paraméterek beállítása
  scale_fill_manual(values = c("#DCDCDC", "#A9A9A9"),
                   label=c(Female="Nő", Male="Férfi")) +
  # a pont (shape) paraméterek beállítása
  scale_shape_manual(values = c(21,22),
                    label=c(Female="Nő", Male="Férfi")) +
  # feliratok hozzáadása
  labs(x = "Dohányzás", y = "Pulzus", fill="Nem", shape="Nem") +
  # fekete-fehér stílus hozzáadása
  theme_bw() +
  # a jelmagyarázat pozíciójának beállítása
  theme(legend.position = "top")

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/atlagabra_07.png", plot = p1,
       units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)
```

```
p1 # a p1 ábra kirajzolása a képernyőre
```



7.2.7 Publikációkész ábra - 4.

Elkészítjük a publikációkész ábrát. Most pontokkal ábrázoljuk az átlagokat. A 95%-os konfidencia intervallumot az `errorbar` segítségével hozzuk létre. Az átlagokat vonallal is összekötjük.

```
# ggplot2 csomagok betöltése
library(ggplot2)

# a p1 objektuma rajzolunk
p1 <- ggplot(data=d, aes(x=Smoke, y=Pulse,
                        fill=Sex, shape=Sex)) +
  # vonal megrajzolása
  stat_summary(fun=mean, geom="line", aes(group=Sex, linetype=Sex),
              colour="black", position=position_dodge(width=0.2)) +
  # hozzá tartozó konfidencia intervallum kirajzolása, errorbar
  stat_summary(fun.data=mean_cl_normal, geom="errorbar", width=0.2,
              position=position_dodge(width=0.2), show_guide = FALSE) +
  # átlagok kirajzolása, pontokkal
  stat_summary(fun=mean, geom="point", position=position_dodge(width=0.2),
              size=2) +
  # az y tengely látható tartományának beállítása
  coord_cartesian(ylim = c(0, 100)) +
  # az x tengely beállítása
```

```

scale_x_discrete(na.translate = FALSE,
                  labels=c(Never="Soha",
                           Occas="Alkalmi",
                           Regul="Rendszeres",
                           Heavy="Erős")) +
# a kitöltés (fill) paraméterek beállítása
scale_fill_manual(values = c("#DCDCDC", "#A9A9A9"),
                  label=c(Female="Nő", Male="Férfi")) +
# a pont (shape) paraméterek beállítása
scale_shape_manual(values = c(21,22),
                   label=c(Female="Nő", Male="Férfi")) +
# a vonal (linetype) paraméterek beállítása
scale_linetype_manual(values = c(1,2),
                     label=c(Female="Nő", Male="Férfi")) +
# feliratok hozzáadása
labs(x = "Dohányzás", y = "Pulzus",
     fill="Nem", shape="Nem", linetype="Nem") +
# fekete-fehér stílus hozzáadása
theme_bw() +
# a jelmagyarázat pozíciójának beállítása
theme(legend.position = "top")

# p1 # a p1 ábra kirajzolása

# a p1 ábra mentése
ggsave("output/kep/atlagabra_08.png", plot = p1,
       units = "in", width = 6, height = 6*0.6, dpi = 300, scale = 0.8)

p1 # a p1 ábra kirajzolása a képernyőre

```

