

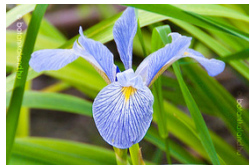
# Ábrakészítés ggplot2 csomaggal

# Iris database

Edgar Anderson: íriszek morfológiai változatosságára gyűjtötte, Ronald Aylmer Fisher (biológus, statisztikus) statisztikai modellek taxonómikus problémáinak feltárására használta.

# Íriszek

"A nőszirmom vagy írisz (Iris) a spárgavirágúak (Asparagales) rendjébe tartozó nőszirmofélék (Iridaceae) családjának névadó nemzetsége. 200-300 faj tartozik ide." (Wiki)



Versicolor, Setosa, Virginica

# A ggplot2 és ábrák készítése általában: ALAPOK I

Az R alapcsomagjában is találhatóak grafikai parancsok, és ennek a felépítése is tanulható. A ggplot2 egy **talán** kicsit kényelmesebb csomag, de leginkább megszokás kérdése.

Kati példái a base csomaggal, tehát az R alapcsomagjával készülnek, mert ezek egyetlen parancssorral gyorsan elkészíthetőek, sok beépített funkcióval. Teklái ggplot2-vel, ezért más lesz a szintaxisuk, szép színesek maguktól. Azt használjátok, amelyik számotokra kényelmesebb. Az alábbiakban a ggplot2-höz adunk ismereteket, de pl itt: <https://towardsdatascience.com/base-plotting-in-r-eb365da06b22> vagy itt <https://intro2r.com/simple-base-r-plots.html> találtok súgót a base-hez.

# A ggplot2 és ábrák készítése általában: ALAPOK II

ggplot2 fő leírása: <https://ggplot2-book.org/index.html>  
De ezen kívül is sok remek minta található egy-egy konkrétabb probléma megoldására.

Ebben az anyagban egy megoldási mintát találtok, de több módon is felépíthető egy ábra. A választott elemek sorrendje nagyjából mindegy.

Vegyük mintául az iris adattáblát. Legyen a `df ← iris`.

A csészelevelek (`sepal`) és a szíromlevelek (`petal`) hossza és szélessége numerikus változó, a fajta (`species`) string.

## A ggplot2 szintaxisa:

Több módon megadható. A leggyakoribb formát mutatjuk, mivel a legtöbb (súgó) ember is ezzel dolgozik.

# A ggplot2 és ábrák készítése általában: ALAPOK III

- **ggplot(df, aes( x = Sepal.Length, y = Sepal.Width, color = Species))**  
⇒ Tehát a df adattáblából a csészelevelek hossza lesz az x-, szélessége az y-tengelyen. Az íriszfajta mentén eltérő színekkel, de ez nem kötelező része! Összességében is ábrázolhatom az adataimat.
- **geom\_point()**  
⇒ Az előbbi sorokból nem derül ki, hogy milyen ábrát szeretnénk, csak hogy milyen elemekből. Egy **geom\_** előtaggal kezdődő paranccsal adhatjuk meg, milyen módon jelenítse meg. Itt pontfelhőt (ún. scatterplotot) kértünk.
  - Ennek a hasáiban is állíthatóak értékek, pl. milyen formát (pl. **shape** = 20), milyen színt (pl. **color** = "darkblue", milyen méretet (pl. **size** = 3) stb. kérek. De csak az állítható, amit az előző paranccsal nem állítottam be. (Itt a color nem állítható tehát, mert azt fentebb állítottam.)
- Innen opcionális formázó elemek következnek. Saját elvásáraid (avagy a megszabott style sheet) alapján elhagyhatod, de használhatod is.

# A ggplot2 és ábrák készítése általában: ALAPOK IV

- **labs**( `x = "Csészelevelek hossza (cm)", y = "Csészelevelek szélessége (cm), color = "Fajta", title = "Íriszek csészelevelei"` )  
⇒ A labs tehát a tengelyfeliratok és a legendcímek beállítását végzi el. Ha üresen hagyod (pl. `color = ""`), akkor nem ír oda semmit. Alapbeállítás: az oszlopok neveit tünteti fel. (title: az ábra címe az ábrán elhelyezve)
- **scale\_color\_manual**(`values = c("darkblue", "brown", "magenta")`)  
⇒ A színek beállítása. Fent kértük, hogy a fajta alapján legyen színezve a sok kis pötty. Ha határozott elképzelésünk van, milyen színűk legyen, megadhatjuk. Több sűgő is van, pl.  
<http://sape.inf.usi.ch/quick-reference/ggplot2/colour>  
Nem kötelező megadni, magától is állít be színt.  
Ha az ábrázolásnál nem az elem kerületét, hanem a területét szeretnénk átszínezni, akkor a parancs: **scale\_fill\_manual**(`values = c(...)`)  
Ha nem kategoriális(ként kezelt) változó alapján színezünk:  
**scale\_x\_continuous**(`name="red", breaks=seq(0.05, 1.05, 0.2), labels=seq(0, 1, 0.2)`)

# A ggplot2 és ábrák készítése általában: ALAPOK V

⇒ Ezt az  $y$ -tengely alapján is megtehetjük. Ebben az ábrában most nincs értelme!

- **theme\_bw()**

⇒ a téma beállítása. Ez általában meghatározza a grafika kinézetét, de az elemei állíthatók tovább.

- **theme(strip.background = element\_rect(fill = "white"),  
strip.text = element\_text(color = "black", size = 16, family = "Times"),  
axis.title = element\_text(color = "black", size = 16, family = "Times"),  
axis.text = element\_text(color = "black", size = 16, family = "Times"),  
legend.text = element\_text(size=16, colour="Black", family = "Times"),  
legend.title = element\_text(size=16, colour="Black", family = "Times", face = "bold"),  
legend.position = "top",  
plot.title = element\_text(size=16, colour="Black", family = "Times"))**



# A ggplot2 és ábrák készítése általában: ALAPOK VI

⇒ A `theme()` paranccsal (nem a `theme_bw()`-ben, hanem külön) állíthatóak az ábra részletei: háttere, karakterek. Azt azonban, hogy a Times a TNR-t tartalmazza, meg kell adni az alábbi paranccsal:

- **`windowsFonts(Times = windowsFont("Times New Roman"))`**

Nem Windowson elvben:

**`library(extrafont)`**

**`font_import()`**

**`fonttable()`**

- `facet_wrap()` és `facet_grid()` ⇒ Az ábra bontása egy vagy két változó alapján.

**`facet_wrap("Species", ncol = 2, nrow = 2)`**

⇒ A `facet_wrap()` egy változó alapján bont. A változó nevét ""-be kell tenni. Az oszlopok (`ncol`) és sorok (`nrow`) számát meg kell adni - legalább az egyiket.

**`facet_grid(.~Species)`**

⇒ Egy vagy két változó alapján bont. A tilde két oldalán állnak a változók. Ha csak egy van, az egyiket pontra kell cserélni (mint a példában). Nincs "" a változók neve körül.

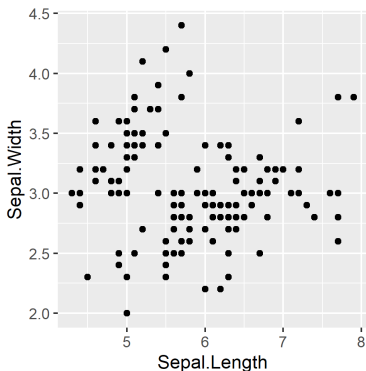
# A ggplot2 és ábrák készítése általában: ALAPOK VII

- Az egyes összetevőket `+`-szal fűzzük össze. Nem kell sort törni sehol, nekem így átláthatóbb.
- Az ábrák többféleképpen menthetők. Pl.:  
**`library(ggpubr)`**  
**`ggsave("...", height = 8, width = 8, units = "cm")`**  
A méretet nem kötelező megadni, akkor az ablak méretében menti el.

# A ggplot2 és ábrák készítése általában: ALAPOK VIII

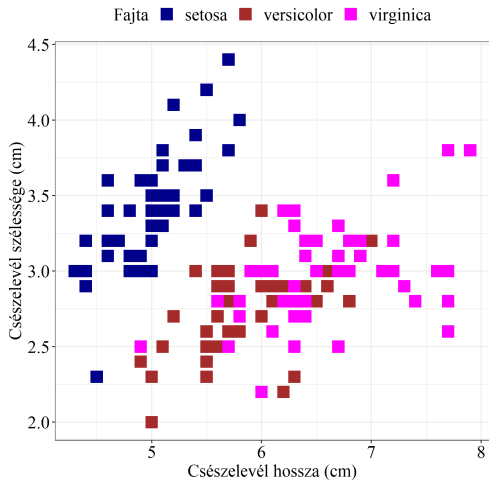
Példa egy sima felhőábra:

```
ggplot(iris, aes(x = Sepal.Length, y = Petal.Length)) +  
geom_point()
```



# A ggplot2 és ábrák készítése általában: ALAPOK IX

Példa egy megformázott felhőábrára:



# A ggplot2 és ábrák készítése általában: ALAPOK X

```
ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width, color = Species))+  
geom_point(size = 5, shape = 15)+  
labs(x = "Csészelevél hossza (cm)", y = "Csészelevél szélessége (cm)",  
color = "Fajta") +  
theme_bw()+  
scale_color_manual(values = c("darkblue", "brown", "magenta")) +  
theme(strip.background = element_rect(fill = "white"),  
strip.text = element_text(color = "black", size = 16, family = "Times"),  
axis.title = element_text(color = "black", size = 16, family = "Times"),  
axis.text = element_text(color = "black", size = 16, family = "Times"),  
legend.text = element_text(size=16, colour="black", family = "Times"),  
legend.title = element_text(size=16, colour="black", family = "Times"),  
legend.position = "top")
```