

Adatelemzés és programozás orvos biológusoknak I.

Sebestyén Endre

Intro – 2019 szeptember 11.

Tematika

- Adatelemzés, programozás, projekt kezelés kezdőknek
- A Carpentries Foundation anyagaira építve
<https://carpentries.org/>
- „The Carpentries teaches foundational coding, and data science skills to researchers worldwide. Software Carpentry, Data Carpentry, and Library Carpentry workshops are based on our lessons.,,

Oktatók

- Cserző Miklós (Élettani Intézet)
- Sebestyén Endre (I. Patológia)
- Szalai Bence (Élettani Intézet)
- Turu Gábor (Élettani Intézet)

Tematika

Szept 4	Sebestyén Endre	Bevezetés, szoftverek installálása
Szept 11	Sebestyén Endre	Adatok rendszerezése, projektek strukturálása, “tidy data”
Szept 18	Sebestyén Endre	Bevezetés az R-be I: R és RStudio, változók, vektorok, listák és funkciók
Szept 25	Sebestyén Endre	Bevezetés az R-be II: részhalmazok képzése, hiányzó adatok kezelése
Okt 2	Sebestyén Endre	Bevezetés az R-be III: adatok beolvasása és kiírása, faktorok, data.frame
Okt 9	Sebestyén Endre	Bevezetés az R-be IV: CRAN és Bioconductor, dplyr használata
Okt 16	Sebestyén Endre	R grafika: adat vizualizálás ggplot2-vel, R Markdown

Tematika

Okt 30	Sebestyén Endre	Linux/UNIX operációs rendszerek és a parancssor R/Rstudio ismétlés, közös problémamegbeszélés
Nov 6	Turu Gábor	Parancssor I: fájlrendszer, fájlok kezelése
Nov 13	Sebestyén Endre	Parancssor II: fájlok manipulálása, shell scriptek
Nov 20	Sebestyén Endre	Verziókövetés: git és github
Nov 27	Szalai Bence	Python I: Bevezetés a Python nyelvbe, Jupyter fejlesztői környezet
Dec 4	Szalai Bence	Python II: Python változó típusok
Dec 11	Szalai Bence	Python III: Elágazások, ciklusok, függvények Python nyelvben

A kurzus kezdőknek szól!



How to actually learn any new programming concept



Essential

Changing Stuff and Seeing What Happens



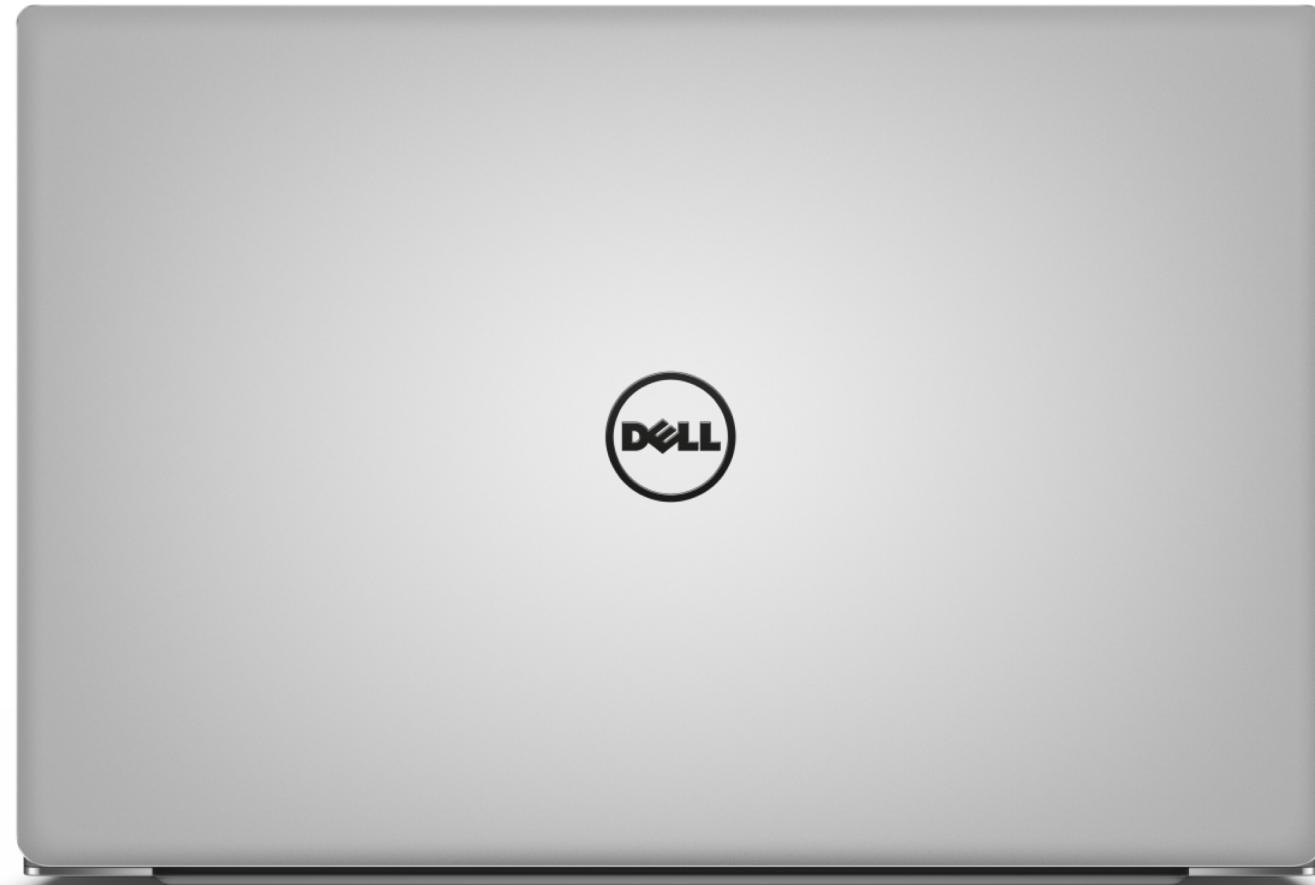
(c) secumem 2009

Code of Conduct

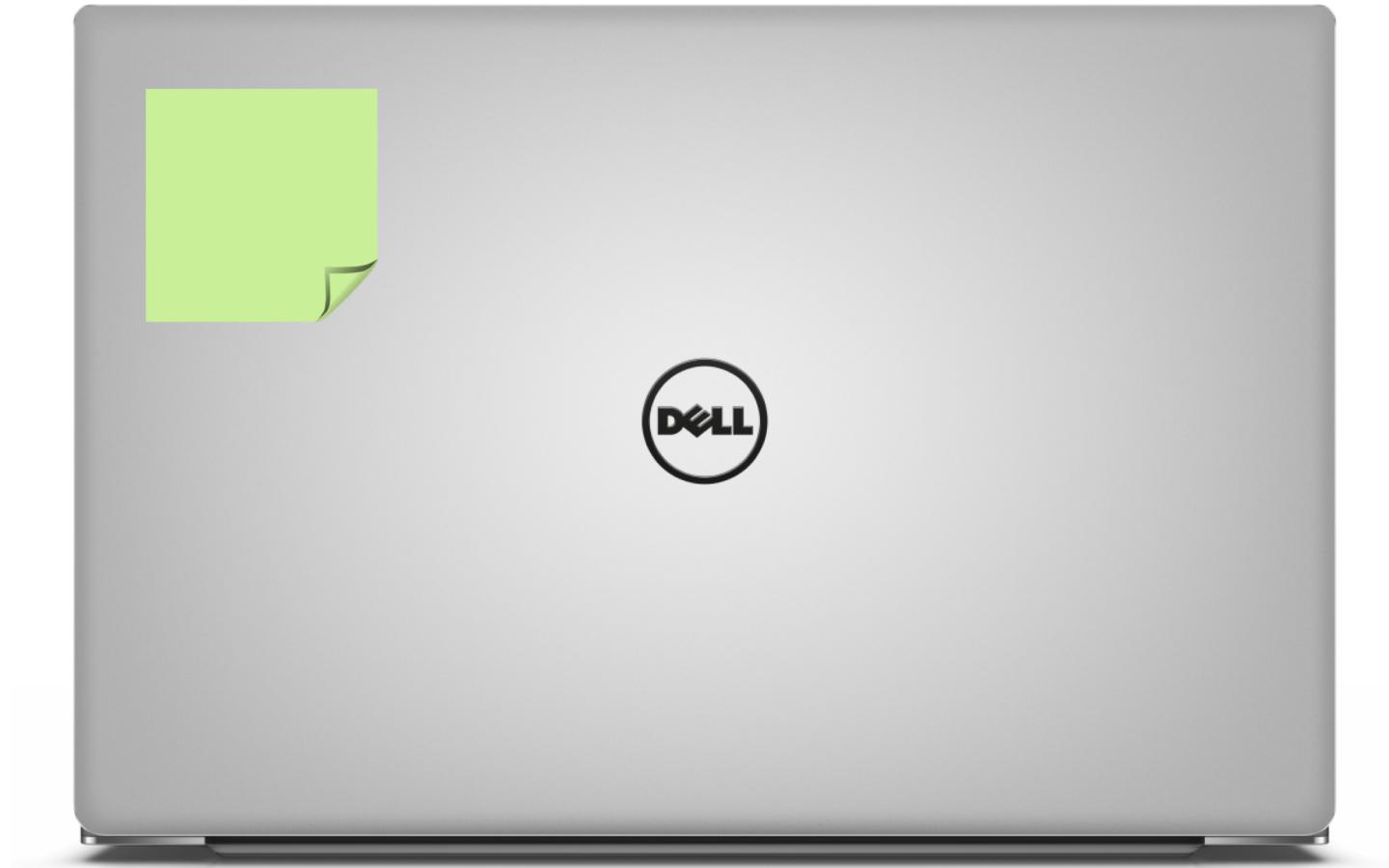
- „The Carpentries is a community-led project. We value the involvement of everyone in the community. We are committed to creating a friendly and respectful place for learning, teaching and contributing. All participants in our events and communications are expected to show respect and courtesy to others.”
- [https://docs.carpentries.org/topic_folders/policies/
code-of-conduct.html](https://docs.carpentries.org/topic_folders/policies/code-of-conduct.html)

HELLO
my name is

A piros/zöld post-it

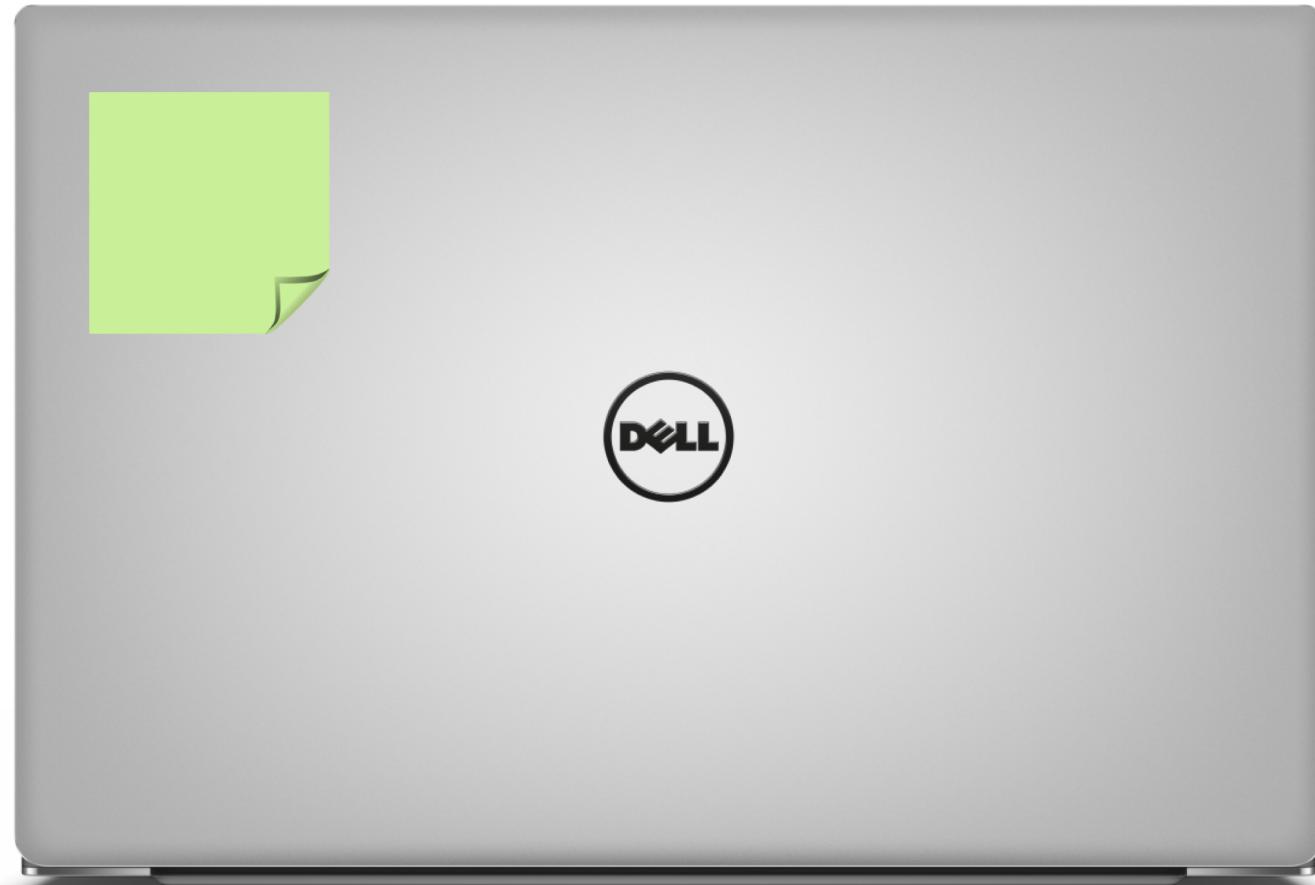


A piros/zöld post-it



A piros/zöld post-it

Minden OK, érhető, nincs kérdés



A piros/zöld post-it



A piros/zöld post-it

Valami nem működik, kérdésem van



Feedback!

- Szept 25, Okt 16, Nov 20, Dec 11
 - A zöld sticky-re valamit ami tetszett
 - A piros sticky-re valamit, amin javítani lehetne
- sebestyen.endre@med.semmelweis-univ.hu
 - Subject: [adatelemzo kurzus]

Tananyag

- <https://su-compbio.github.io/course/>
 - Menetrend
 - Prezentációk
 - Összefoglaló
 - Ajánlott cikkek
 - Linkek
 - stb

Házi feladatok

- 12 házi feladat
- < 6 😞
- 6+ 😊
- 10+ 😎
- Dropbox File Request feltöltés vasárnap éjfélig
- ***A lényeg: hogyan próbáltad megoldani, mi működik, mi nem működik***

- Hogy hívnak és hol kutatsz?
- Eddigi tapasztalat programozással, adatelemzéssel?
- Valami, amit a múlt héten csináltál es nagyon büszke vagy rá, nagyon örültél neki



Adatrendszerzés excelben (és hasonló szoftverekben)

The screenshot shows two Excel tabs side-by-side. The left tab, titled 'survey_data.xlsx', contains raw survey data with columns: DateCollected, Year, Month, Day, Plot, Species, Sex, and Weight. The data spans from row 1 to 11, with rows 1-2 as headers. The right tab, titled 'README_surveyData.txt', contains processing notes. The notes describe the steps taken to clean the data, including transferring raw data to clean versions for 2013 and 2014, creating a 'Species' column, and separating the date into year, month, and day. The notes also mention moving information from the header to the 'Species' column.

	A	B	C	D	E	F	G	H	I
1	DateCollected	Year	Month	Day	Plot	Species	Sex	Weight	
2	7/16/13	2013	7	16	2	DM	F		
3	7/16/13	2013	7	16	7	DM	M	33g	
4	7/16/13	2013	7	16	3	DM	M		
5	7/16/13	2013	7	16	1	DM	M		
6	7/18/13	2013	7	18	3	DM	M	40g	
7	7/18/13	2013	7	18	7	DM	M	48g	
8	7/18/13	2013	7	18	4	DM	F	29g	
9	7/18/13	2013	7	18	4	DM	F	46g	
10	7/18/13	2013	7	18	7	DM	M	36g	
11	7/18/13	2013	7	18	7	DM	F	35g	

Processing notes on survey_data.xlsx

2014-08-19 work done -----

1. Transferred 2013-raw to 2013-clean, and 2014-raw to 2014-clean
2. In 2013-clean: created a 'Species' column and moved information from header to that column
3. In 2013-clean, put all the different tables together into one table with columns: date collected, plot, species, sex, weight
4. In 2013-clean, separated month/day/year column into three columns for year, month, and day using formulas

Példa adatsor: jó ez ebben a formában?

Date collected	Plot	Species-Sex	Weight
1/9/78	1	DM-M	40
1/9/78	1	DM-F	36
1/9/78	1	DS-F	135
1/20/78	1	DM-F	39
1/20/78	2	DM-M	43
1/20/78	2	DS-F	144
3/13/78	2	DM-F	51
3/13/78	2	DM-F	44
3/13/78	2	DS-F	146

Példa adatsor: Species/Sex külön oszlopba!

Date collected	Plot	Species	Sex	Weight
1/9/78	1	DM	M	40
1/9/78	1	DM	F	36
1/9/78	1	DS	F	135
1/20/78	1	DM	F	39
1/20/78	2	DM	M	43
1/20/78	2	DS	F	144
3/13/78	2	DM	F	51
3/13/78	2	DM	F	44
3/13/78	2	DS	F	146

Dátumok

Milyen adatokat terveztek elemezni/gyűjteni?

