Kako organizirati podatke u istraživačkom projektu?

Denis Vlašiček
Odsjek za psihologiju FFZG | CROSSDA
dvlasice@ffzg.hr
@dvlasicek
vdeni

1) Uvod

2) Organizacija datoteka

3) Upravljanje istraživačkim podacima

4) Praćenje promjena

5) Dodatni materijali i reference

1) Uvod

Uvod

- Što znači organizirati podatke u istraživačkom projektu?

Uvod

- Što znači organizirati podatke u istraživačkom projektu?
- Zbog čega treba razmišljati o svemu tome?

Uvod

- Što znači organizirati podatke u istraživačkom projektu?
- Zbog čega treba razmišljati o svemu tome?
- Kakve veze to ima s *istraživanjima*?

- Ali oduzima puno vremena...

- Ali oduzima puno vremena...
  - ▷ vraćanje projektu nakon bavljenja nečim drugim

- Ali oduzima puno vremena...
  - ▷ vraćanje projektu nakon bavljenja nečim drugim
  - ▷ automatizacija (dijela) procesa

- Ali oduzima puno vremena...
  - ▷ vraćanje projektu nakon bavljenja nečim drugim
  - ▷ automatizacija (dijela) procesa
  - ▷ ponovna uporaba materijala

- Ali oduzima puno vremena...
  - ▷ vraćanje projektu nakon bavljenja nečim drugim
  - ▷ automatizacija (dijela) procesa
  - ▷ ponovna uporaba materijala
  - ▷ lakše je reproducirati nalaze

2) Organizacija datoteka

Organizacija datoteka

- svaki projekt treba dobiti vlastiti folder
  - ▷ kako definirati što je *projekt*

Organizacija datoteka

- svaki projekt treba dobiti vlastiti folder
  - kako definirati što je *projekt*

- napraviti posebne foldere za različite tipove datoteka
  - podaci (`data`)
  - tekstualne datoteke (rukopisi, dokumentacije) (`doc`)
  - datoteke vezane uz analize (`stats`)

Organizacija datoteka

- svaki projekt treba dobiti vlastiti folder
  - kako definirati što je *projekt*
- napraviti posebne foldere za različite tipove datoteka
  - podaci (`data`)
  - tekstualne datoteke (rukopisi, dokumentacije) (`doc`)
  - datoteke vezane uz analize (`stats`)
- imena datoteka trebaju odražavati njihovu funkciju

Organizacija datoteka

- svaki projekt treba dobiti vlastiti folder
  - ▷ kako definirati što je *projekt*
- napraviti posebne foldere za različite tipove datoteka
  - ▷ podaci (`data`)
  - ▷ tekstualne datoteke (rukopisi, dokumentacije) (`doc`)
  - ▷ datoteke vezane uz analize (`stats`)
- imena datoteka trebaju odražavati njihovu funkciju

```
megahrt-zadar/
├── analyses
│   ├── data
│   │   ├── psycholing-estimates
│   │   │   ├── age-of-acquisition.csv
│   │   │   ├── concreteness.csv
│   │   │   ├── raw-estimates
│   │   │   │   ├── dob.txt
│   │   │   │   ├── konkretnost.txt
│   │   │   │   └── subjektivna.txt
│   │   │   └── subjective-frequency.csv
│   │   ├── reaction-time
│   │   │   ├── clean-non-printable.sh
│   │   │   ├── megahrt-part1-clean.csv
│   │   │   ├── megahrt-part1.csv
│   │   │   ├── megahrt-part2-clean.csv
│   │   │   └── megahrt-part2.csv
│   │   └── words
│   │       ├── words-wave-1.csv
│   │       └── words-wave-2.csv
│   ├── helpers
│   │   ├── 01-psycholing-evaluations.R
│   │   └── 02-plot-helpers.R
│   ├── plots
│   │   ├── 01-plot-ppc-overall.R
│   │   ├── 02-plot-ppc-participant.R
│   │   ├── 03-effects-plots.R
│   │   ├── effects-plot.png
│   │   ├── ppc-overall.png
│   │   └── ppc-per-participant.png
│   ├── stats
│   │   ├── 01-prepare-data-rt-analysis.R
│   │   ├── 02-rt-model.stan
│   │   ├── 03-model-fitting.R
│   │   └── 04-ppc-sampling.R
│   └── wrangling
│       ├── 01-rt-data-anonimization.R
│       │       ...
│       └── 05-combine-datasets.R
├── presentation
│   ├── hrzz-logo.jpg
│   ├── megahr-logo.png
│   ├── megahr-zadar.pdf
│   └── megahr-zadar.tex
├── README.md
└── reference.bib
```

3) Upravljanje istraživačkim podacima

Upravljanje istraživačkim podacima

- spremite sirove podatke!
  - pohranite ih na sigurno mjesto
  - onemogućite izmjenu podataka (read-only)
  - nemojte ih zamijeniti pročišćenim podacima

Upravljanje istraživačkim podacima

- spremite sirove podatke!
  - pohranite ih na sigurno mjesto
  - onemogućite izmjenu podataka (read-only)
  - nemojte ih zamijeniti pročišćenim podacima

- napravite sigurnosne kopije, i pohranite ih na nekoliko različitih mjesta

Upravljanje istraživačkim podacima

- spremite sirove podatke!
  - pohranite ih na sigurno mjesto
  - onemogućite izmjenu podataka (read-only)
  - nemojte ih zamijeniti pročišćenim podacima
- napravite sigurnosne kopije, i pohranite ih na nekoliko različitih mjesta
- ako možete, koristite i otvorene formate za pohranu
  - primjerice `CSV` (comma separated values) uz `SAV`

- napravite uredne (*tidy*) podatke
  - svaki stupac je jedna varijabla (ne spajati dvije vrijednosti u isti stupac)
  - svaki red je jedno opažanje/jedinica analize
  - ne ostavljajte prazne ćelije

- napravite uredne (*tidy*) podatke
  - ▷ svaki stupac je jedna varijabla (ne spajati dvije vrijednosti u isti stupac)
  - ▷ svaki red je jedno opažanje/jedinica analize
  - ▷ ne ostavljajte prazne ćelije

NAZIV	SKRACENI_NAZIV	OSOBE	OIB	STATUS
MEĐIMURSKA STRANKA	MS	Grašić Damir - PREDSJEDNIK - predstavlja i zastupa	36176331004	AKTIVNA
DEMOKRATSKA STRANKA UMIROVLJENIKA	DSU	Tičarić Milivoj - PREDSJEDNIK - predstavlja i zastupa	19578261879	AKTIVNA
ZELENA STRANKA	ZS	Babić Aljoša - PREDSJEDNIK - predstavlja i zastupa	40857015687	AKTIVNA
JADRANSKA SOCIJALDEMOKRATSKA STRANKA HRVATSKE	JADRANSKI SOCIJALDEMOKRATI	Kovačić Boris - PREDSJEDNIK - predstavlja i zastupa; Marušić Josip - DOPREDSJEDNIK - predstavlja i zastupa	92762294306	AKTIVNA
DEMOKRATSKA KNEGINEČKA STRANKA	DKS	Keretić Damir - PREDSJEDNIK - predstavlja i zastupa	86378509070	AKTIVNA
AUTONOMNA REGIONALNA STRANKA HRVATSKOG PRIMORJA,	ARS	Jelušić Ivo - PREDSJEDNIK - predstavlja i zastupa	96925777632	AKTIVNA
AUTOHTONA - HRVATSKA STRANKA PRAVA	A - HSP	Keleminec Dražen - PREDSJEDNIK - predstavlja i zastupa	84404674327	AKTIVNA
ZELENI SAVEZ	ZELENI	Rupnik Josip Anton - PREDSJEDNIK - predstavlja i zastupa	58846771856	AKTIVNA
HRVATSKA STRANKA NEZAPOSLENIH	HSN	Milinović Viktor - PREDSJEDNIK - predstavlja i zastupa	82664540044	AKTIVNA
HRVATSKA RADNIČKA STRANKA	HRS	Novosel Mladen - PREDSJEDNIK - predstavlja i zastupa	36281215365	AKTIVNA
ZAGORSKA STRANKA	ZS	Jernečić Miljenko - predstavlja i zastupa	80505604089	AKTIVNA
DEMOKRATSKA STRANKA ŽENA	DSŽ	Jelinčić Marija - PREDSJEDNIK - predstavlja i zastupa	27223591444	AKTIVNA
SOCIJALISTIČKA PARTIJA HRVATSKE	SPH	Šipov Dejan - PREDSJEDNIK - predstavlja i zastupa	18440010529	AKTIVNA
AKCIJA MLADIH	AM	Rubeša Vjekoslav - PREDSJEDNIK - predstavlja i zastupa; Buković Darjan - POTPREDSJEDNIK - predstavlja i zastupa;+	17604594468	AKTIVNA
MEĐIMURSKI DEMOKRATSKI SAVEZ	MDS	Pavlic Željko - PREDSJEDNIK - predstavlja i zastupa	22716317556	AKTIVNA
ALTERNATIVA		Simonić Bojan - PREDSJEDNIK - predstavlja i zastupa		AKTIVNA
HRVATSKI DEMOKRATSKI SAVEZ SLAVONIJE I BARANJE	HDSSB	Glavaš Branimir - PREDSJEDNIK - predstavlja i zastupa	37050651240	AKTIVNA
LISTA ZA RIJEKU	RI	Švorinić Danko - PREDSJEDNIK - predstavlja i zastupa; Basic Sandi - DOPREDSJEDNIK - predstavlja i zastupa	3376628315	AKTIVNA
HRVATSKA SELJAČKA STRANKA BRAĆE RADIĆ	HSS BRAĆE RADIĆ	Grčić Stanko - PREDSJEDNIK - predstavlja i zastupa	36114066915	AKTIVNA
HRVATSKA STRANKA PRAVNE DRŽAVE	HSPD	Prpić Siniša - PREDSJEDNIK - predstavlja i zastupa	56694372533	AKTIVNA
JEDINO HRVATSKA - POKRET ZA HRVATSKU	JEDINO HRVATSKA	Šibl Milovan - PREDSJEDNIK - predstavlja i zastupa	43647783176	AKTIVNA
SPLITSKA STRANKA UMIROVLJENIKA	SSU	Kovačić Ratko - PREDSJEDNIK - predstavlja i zastupa	18699517883	AKTIVNA
BOŠNJAČKA DEMOKRATSKA STRANKA HRVATSKE	BDSH	Hodžić Nedžad - PREDSJEDNIK - predstavlja i zastupa	39202199526	AKTIVNA
POPRAVI GRAD - NOVA GENERACIJA	PG - NG	Zoričić Ante - PREDSJEDNIK - predstavlja i zastupa	90654406562	AKTIVNA
NOVA SRPSKA STRANKA	NSS	Ladarević Svetislav - PREDSJEDNIK - predstavlja i zastupa	77799192134	AKTIVNA
DEMOKRATSKA PARTIJA SRBA	DPS	Kojadinović Mitar - PREDSJEDNIK - predstavlja i zastupa	40735782120	AKTIVNA
PAŠKA STRANKA - PAG	PSP	Marčić Željko - PREDSJEDNIK - predstavlja i zastupa	983779256	AKTIVNA

- bilježite sve korake procesiranja skupa podataka

- bilježite sve korake procesiranja skupa podataka
  - ▷ procesiranje podataka je jedan od ključnih koraka u analizi

- bilježite sve korake procesiranja skupa podataka
  - ▷ procesiranje podataka je jedan od ključnih koraka u analizi
  - ▷ mora biti moguće pratiti postupak od sirovih podataka do podataka korištenih u obradama

- bilježite sve korake procesiranja skupa podataka
  - ▷ procesiranje podataka je jedan od ključnih koraka u analizi
  - ▷ mora biti moguće pratiti postupak od sirovih podataka do podataka korištenih u obradama

- napravite dokumentaciju za podatke (*metapodatke*)

4) Praćenje promjena

Praćenje promjena

- olakšava praćenje istraživačkog procesa
  - kad je nastala promjena u nekom dokumentu
  - vraćanje na starije verzije

Praćenje promjena

- olakšava praćenje istraživačkog procesa
  - kad je nastala promjena u nekom dokumentu
  - vraćanje na starije verzije
- olakšava reproduciranje postupaka koji su doveli do nekog ishoda (npr. istraživački rad, izvještaj)

Praćenje promjena

- olakšava praćenje istraživačkog procesa
  - kad je nastala promjena u nekom dokumentu
  - vraćanje na starije verzije
- olakšava reproduciranje postupaka koji su doveli do nekog ishoda (npr. istraživački rad, izvještaj)
- specijalizirani alati (`git`) i servisi (GitHub) ili ručno

- u folder projekta dodati `CHANGELOG.txt` dokument
  - ▷ bilješke s izmjenama učinjenima na datotekama projekta
  - ▷ svaki unos ima vremensku oznaku, najnoviji unosi na početku

- u folder projekta dodati `CHANGELOG.txt` dokument
  - ▷ bilješke s izmjenama učinjenima na datotekama projekta
  - ▷ svaki unos ima vremensku oznaku, najnoviji unosi na početku

```
#### [2020-11-05_12-00]
- prosirio primjer za rucno pracenje izmjena
- dodao slike
#### [2020-11-05_11-15]
- dodao primjer za rucno pracenje izmjena
```

- napravite kopiju projekta kad god učinite neku značajnu promjenu

- napravite kopiju projekta kad god učinite neku značajnu promjenu
- kopiju pohranite u folder unutar projektnog foldera
- kao ime foldera s kopijom stavite datum kad je kopija napravljena

- napravite kopiju projekta kad god učinite neku značajnu promjenu
- kopiju pohranite u folder unutar projektnog foldera
- kao ime foldera s kopijom stavite datum kad je kopija napravljena

```
projekt/
|-- trenutna
|-- 2020-11-05
|-- 2020-11-04
```

- napravite kopiju projekta kad god učinite neku značajnu promjenu
- kopiju pohranite u folder unutar projektnog foldera
- kao ime foldera s kopijom stavite datum kad je kopija napravljena

```
projekt/
|-- trenutna
|-- 2020-11-05
|-- 2020-11-04
```

- mediji za pohranu su jeftini, podaci su skupi!

- Puh, Dropbox i Google Drive omogućavaju pohranu različitih verzija datoteka

- Google Docs omogućava pohranu različitih verzija tekstova

?

?

- Kako organizirate svoje projekte?
- Kako dijelite materijale sa suradnicima i kolegama?
- Gdje pohranjujete podatke za vrijeme trajanja projekta?

Dodatni materijali i reference

- morževi preuzeti iz Kastelein, R. A., Zweypfenning, R. C. V. J., Spekreijse, H., Dubbeldam, J. L., i Born, E. W. (1993). The Anatomy of the walrus head (Odobenus rosmarus). Part 3: The eyes and their function in Walrus ecology. *Aquatic mammals, 19*, 61-61.
- sadržaj temeljen na Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., i Teal, T. K. (2017). Good enough practices in scientific computing. *PLoS computational biology, 13*(6), e1005510. doi: `https://doi.org/10.1371/journal.pcbi.1005510`
- dodatna inspiracija iz CESSDA Training Team (2020). *CESSDA Data Management Expert Guide*. Bergen, Norway: CESSDA ERIC. doi: `https://doi.org/10.5281/zenodo.3820472`
- sustav za pohranu podataka `puh` (Srce): `https://www.srce.unizg.hr/puh`
- kako imenovati datoteke: `https://speakerdeck.com/jennybc/how-to-name-files`