

# Statistiline andmeteadus ja visualiseerimine

[andmeteadus.github.io/2017](https://andmeteadus.github.io/2017)

Taavi Unt ja Oliver Aasmets

[taavi.unt@ut.ee](mailto:taavi.unt@ut.ee), [oliver.aasmets@gmail.com](mailto:oliver.aasmets@gmail.com)

# Õppetöö korraldus

- praktikumid
  - kokku 9 tk
  - reedeti 14:15 ruumis 203
  - küllalt soovituslik teha oma laptopiga
- loenguid ei ole
- kodutööd
  - kokku 8 tk
- projekt

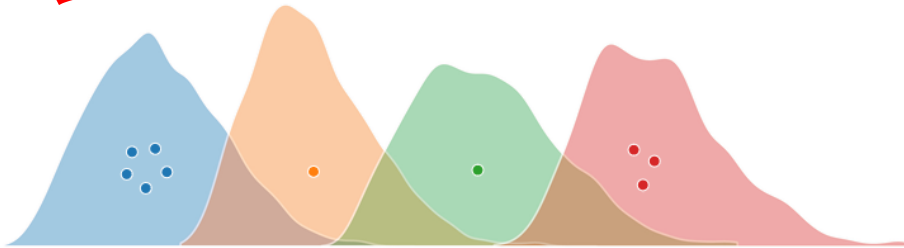
# Aine läbimiseks on vaja

1. **Saada vähemalt 150 punkti** kodutööde eest.
  - Iga kodutöö kuni 25 punkti, kokku kuni  $8 \times 25 = 200$  punkti.
  - Kodutöödes on **boonus**ülesandeid, mis annavad lisapunkte.
2. **Teha projekt**, kirjutada populaarteaduslikus stiilis ülevaade / teha Shiny rakendus. Saadud tulemusi suuliselt esitleda.
  - Kas üksinda või kahepeale.
  - Teema valik: otsige ise või oodake, kuni meie midagi pakume.
  - Teemad, tähtajad jms ilmuvad lehele <http://andmeteadus.github.io/2017/projektid/>

# Kodutööd

Statistiline andmeteadus ja visualiseerimine MTMS.01.100

Praktikumid Projektid



## Praktikumimaterjalid

### Praktikum 1

10.02.2015

Sissejuhatus ainesse, reprodutseeritavasse analüüsi ja automatiseeritud statistikasse. Tarkvara R meeldetuletamine. Pakett knitr. Markdown. Pakett dplyr. Näide: automaatne aktsiaturu ülevaade.

## Õppeaine lühikirjeldus

Tegemist on rakendusliku statistikakursusega, kus:

- Probleemipüstitusena kasutatakse huvitavaid päriselulisi andmestikke.
- Rakendatakse sobivaid statistilisi meetodeid, et püstitatud küsimustele vastust leida.
- Saadud tulemuste teistele kommunikeerimiseks illustreeritakse neid sobivate joonistega.

Järgnev skeem võtab lühidalt kokku õppeaine olemuse.

- Kursuse “Rakendustarkvara: R” õppematerjal.
- Praktikumis kasutatud materjalid:
  - Slaidid (korralduslik info)
  - Logige *Socratic* tagasiside keskkonda koodiga 5fc6f3ec
  - knitr ja R Markdown sissejuhatus
  - dplyr minimalistlik näide
- Automaatne aktsiaturu ülevaade
- Kodutöö
  - Tähtaeg: 16.02.2015 kell 23.59 GMT

• tähtaeg järgneval neljapäeval kell 23:59

• saata aadressile [taavi.unt@ut.ee](mailto:taavi.unt@ut.ee)

# Projekt

- Üksinda või kahepeale tuleb analüüsida ja visualiseerida.
  - Teema valimise tähtaeg on [07.04.2017](#).
  - Teema (s.t. andmed) võib otsida ise või oodata meie pakkumist (tulevad 31.03.2017).
  - Soovitame ise otsimisega tegeleda alles pärast 5. praktikumi (andmete kraapimine veebist, 17.03.2017).
  - Projekti [tähtaeg 28.04.2017](#) ja [ettekanded 05.05.2017](#).
- 2015. aastal tehtud projektid <http://andmeteadus.github.io/2015/projektid/> .
- Suur ideaal, mille poole projektides püüelda: [www.stat24.ee](http://www.stat24.ee) .
- Projekti juhend [http://andmeteadus.github.io/2017/projekt\\_juhend/](http://andmeteadus.github.io/2017/projekt_juhend/)

# Aeg

- 3 EAP = 78 h jagu töömahtu jaguneb
  - 8 x 1,5 h = 12 h praktikumid
  - 8 x 5 h = 40 h kodutööd
  - 24,5 h projekti koostamine
  - 1,5 h projekti esitlemine (ja teiste esitluste kuulamine)

# Muud põnevad

- [Hadley Wickham](#)

- Hadley.nz
- R for Data Science <http://r4ds.had.co.nz/>

- [datacamp.com](#)

- data.table ja fread: <https://www.datacamp.com/courses/data-analysis-the-data-table-way>

- [Interaktiivsus](#) ja [dashboard](#)

- <http://www.htmlwidgets.org/index.html>

- [useR](#) konverents

- <https://channel9.msdn.com/Events/useR-international-R-User-conference/useR2016>
- <http://www.user2017.brussels/>

# MTMS.01.100

teoreetilised  
statistika-  
teadmised

tegelikkuses  
vajalikud  
oskused