

# Statistiline andmeteadus ja visualiseerimine

[andmeteadus.github.io/2019](https://andmeteadus.github.io/2019)

Kaur Lumiste

[kaur.lumiste@ut.ee](mailto:kaur.lumiste@ut.ee)

# Õppetöö korraldus

- praktikumid
  - kokku 9 tk
  - esmaspäeviti 10:15 ruumis 004
  - küllalt soovituslik teha oma laptopiga
- loenguid ei ole
- kodutööd
  - kokku 8 tk
- projekt

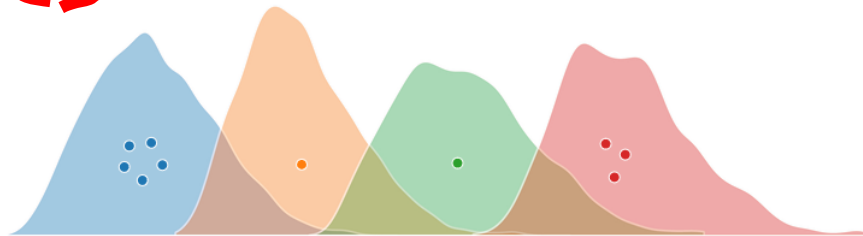
# Aine läbimiseks on vaja

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/2018/projektid/>

# Kodutööd

## Statistiline andmeteandus ja visualiseerimine MTMS.01.100

Praktikumid Projektid



### 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.

## Praktikum 1

14.02.2018

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

- [Kursuse "Rakendustarkvara: R" õppematerjal.](#)
- Praktikumis kasutatud materjalid:
  - [Slaidid](#) (korralduslik info)
  - Sisenege [Socrative](#) keskkonda
  - [knitr ja R Markdown sissejuhatus](#)
  - [dplyr minimalistlik näide](#)
- [Automaatne aktsiaturu ülevaade](#)
- [Kodutöö](#)
  - Tähtaeg 20.02.2018 kell 23:59.

- tähtaeg järgneval [esmaspäeval kell 10:15](#)
  - st 1 nädal aega, et saada kõik väljateenitud punktid
  - kuni 1 nädal hilinemist (ülejärgnev esmaspäev kell 10:15) – 50% punktidest
  - pärast seda 0%
- Kodutööd esitada [Moodle keskkonnas](#). Moodle's on näha tulemused.

# Projekt

- Üksinda või kahepeale tuleb analüüsida ja visualiseerida.
  - Teema valimise tähtaeg on [25.03.2019](#).
  - Teema (s.t. andmed) võib otsida ise või valida mõni minu pakkumistest. Soovitan ise otsimisega tegeleda alles pärast 5. praktikumi (andmete kraapimine veebist, 11.03.2018).
  - Projekti [tähtaeg 12.05.2019](#) ja [ettekanded 20.05.2019](#).
- 2015. aastal tehtud projektid <http://andmeteadus.github.io/2015/projektid/>
- 2017. aastal tehtud projektid <http://andmeteadus.github.io/2017/projektid/>
- 2018. aastal tehtud projektid <http://andmeteadus.github.io/2018/projektid/>
- Suur ideaal, mille poole projektides püüelda: [www.stat24.ee](http://www.stat24.ee)
- Projekti juhend [http://andmeteadus.github.io/2019/projekt\\_juhend/](http://andmeteadus.github.io/2019/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](#)
  - [www.htmlwidgets.org/index.html](http://www.htmlwidgets.org/index.html)
- [www.r-bloggers.com](#)
- [useR konverents](#)
  - [channel9.msdn.com/Events/useR-international-R-User-conferences/useR-International-R-User-2017-Conference](http://channel9.msdn.com/Events/useR-international-R-User-conferences/useR-International-R-User-2017-Conference)
  - [user2018.r-project.org](http://user2018.r-project.org) (Brisbane)

# MTMS.01.100

teoreetilised  
statistika-  
teadmised

tegelikkuses  
vajalikud  
oskused