

# Duomenų analizės įvadas

## Intro

Justas Mundeikis

2019 m. vasario 26 d.

# In God we trust, all others bring data

- W. Edwards Deming

# 1 lentelė: Shared 'Dublin' descriptors for Short Cycle, First Cycle, Second Cycle and Third Cycle Awards, 2004

	Bachelor	Master	Doctorate
Knowledge and understanding:	[Is] supported by advanced text books [with] some aspects informed by knowledge at the forefront of their field of study ..	provides a basis or opportunity for originality in developing or applying ideas often in a research* context ..	[includes] a systematic understanding of their field of study and mastery of the methods of research* associated with that field..
Applying knowledge and understanding:	[through] devising and sustaining arguments	[through] problem solving abilities [applied] in new or unfamiliar environments within broader (or multidisciplinary) contexts ..	[is demonstrated by the] ability to conceive, design, implement and adapt a substantial process of research* with scholarly integrity .. [is in the context of] a contribution that extends the frontier of knowledge by developing a substantial body of work some of which merits national or international refereed publication .
Making judgements:	[involves] gathering and interpreting relevant data ..	[demonstrates] the ability to integrate knowledge and handle complexity, and formulate judgements with incomplete data ..	[requires being] capable of critical analysis, evaluation and synthesis of new and complex ideas..
Communication	[of] information, ideas, problems and solutions ..	[of] their conclusions and the underpinning knowledge and rationale (restricted scope) to specialist and non-specialist audiences (monologue) ..	with their peers, the larger scholarly community and with society in general (dialogue) about their areas of expertise (broad scope)..
Learning skills ..	have developed those skills needed to study further with a high level of autonomy ..	study in a manner that may be largely self-directed or autonomous..	expected to be able to promote, within academic and professional contexts, technological, social or cultural advancement ..

# Kurso tikslas

- Suteikti pagrindines reikalingas kompetencijas darbui su duomenimis
- Ugdyti studentų gebėjimą savarankiškai rinkti ir analizuoti mokslinę literatūrą bei duomenis, pateikti apibendrintas išvalgas
- Kursas yra "lengvas", tačiau reikalauja daug praktinio darbo pastangų
- Dirbsime tik su R, RStudio, Git ir Github (Jupyter?)
- Neliesime Matlab, Octave, Python, Julia, Eviews

# Kurso turinys

- 1 Command-line interface, Git ir GitHub, Google Scholar
- 2 R ir RStudio, R programavimo pagrindai
- 3 Tiriamoji duomenų analizė (angl.: exploratory data analysis)
- 4 Atkartojami tyrimai (angl.: reproducible research) su LaTeX, RMarkdown

# Sando pristatymas

- Kurso apimtis 130 akad. val.
  - Paskaitos 32 akad. val.
  - Seminarai 16 akad. val.
  - Savarankiškas darbas 82 akad. val.
- 6 akad. val. per 2 savaites
- 10.5 akad. val. savarankiško darbo per dvi savaites

# Studijos forma

- Paskaitos
  - Paskaitos vyks prie PC
  - Idealiu atveju po 2 paskaitų, trečia paskaita bus seminaras, kai studentai savarankiškai bandys spręsti uždavinius
  - Taip pat paskaitų ir seminarų metu bus aptariamos namų darbų užduotys
  - Paskaitos ir seminarai nėra privalomi, bet rekomenduotini
- Savarankiškas darbas
  - Namų darbų ruošimas (laiko langas uždaromas 2 val. prieš paskaitą), du bandymai su 6 val. pertrauka
  - Rašto darbo paruošimas (esė)
  - Kitų ( $\approx 2 - 3$ ) studentų rašto darbų vertinimas
  - Savarankiškas pateiktos literatūros studijavimas
  - Žinių skaitymas

- Visa studijų medžiaga [Github](#), tik pirmos paskaitos medžiaga @VMA
- Kontaktai: justas.mundeikis@evaf.vu.lt
  - Subject: "data-science"
  - Atsakau įprastai per 3 dd.



# Kurso vertinimo strategija

- Namų darbai - 20%

Namų darbų tikslas parodyti studentams, kokio pobūdžio uždavinius studentai turi gebėti savarankiškai spręsti. Namų darbų vertinimas: maksimumas iš leidžiamų 2 bandymų, arba grupinis vertinimas

- Neanonsuoti testai - 20%

Neanonsuoti trumpi testai skirti užtikrinti, jog studentai nuolatos skirtų deramą dėmesį ir laiku įdėtų reikalingas pastangas studijoms

- Savarankiškas darbas (esė) - 20%

- Baigiamasis egzaminas - 40%

Egzamine tikrinamos tiek teorijos tiek įgytos praktinės žinios  
Egzaminas turi būti išlaikytas ne mažiau 50%, jeigu rezultatas  $< 50\%$  egzmainas turi būti perlaikytas

- $BP = \text{mean}(0.2 * \text{mean}(ND) + 0.2 * \text{mean}(NT) + 0.2 * SD + 0.4 * BE)$

# Tvarka egzamino metu

- Studentas neturi naudotis jokiais šaltiniais ir priemonėmis, kad nesukeltų įtarimų dėl savo nesąžiningumo studijų rezultatų vertinimo metu.
- Neleistinų šaltinių ir priemonių turėjimas, interneto tinklalapių atsidarymas pripažįstamas pakankamu įrodymu, kad studentas šiomis priemonėmis naudojosi.
- Identifikavus, kad studentas naudojasi neleistinais šaltiniais ar priemonėmis, jam toliau egzamino laikyti neleidžiama, o informacija apie įvykį perduodama fakulteto administracijai.
- Studentai / stebėtojai prižiūrės tvarką egzamino metu, bus įrašomas video ekranų turinys
- Konkrečiai kas galima, kas ne, nuspręsimė semestro pabaigoje

# Esė darbas (bus papildyta...!)

- Tema iki 2019-04-01 (sugalvoti savarankiškai)
- Planas iki 2019-05-01
- Įkėlimas iki 2019-06-05
- Apimtis:  $\geq 3000$  žodžių (10-15 psl)
- Kalba: LT/EN
- Vertinimo kriterijai:
  - Esė turinio ir temos pavadinimo atitikimas
  - struktūros aiškumas
  - loginė dėstymo seka, gebėjimas sieti mintis
  - argumentų tinkamumas, teiginių logika ir aiškumas;
  - originalumas;
  - išvadų pagrįstumas
  - citavimo teisingumas ir įforminimas
  - veikiantis R kodas
  - Mažiausiai 10 mokslinių šaltinių finaliniame variante

# Esė planas

Esė darbo planą turėtų sudaryti:

- Įvadas, kuriame nurodomas esė tikslas (pagrindinis klausimas), tarpiniai uždaviniai, esė objektas
- 3-5 pagrindinės dalys
- Esė struktūra, kuri turi turėti aiškią sistemą ir tos sistemos logiką
- Trumpai aprašyta kokie klausimai bus aptariami kiekvienoje dalyje, kaip ir kodėl.
- Literatūros sąrašas, naudotas planui parengti mažiausiai 5 moksliniai šaltiniai

# Literatūra

- Nurodyta sandė
- Tačiau praktiškai, kiekvieno skaidrių set'o pabaigoje, bus pateikti rekomenduotini šaltiniai

# Pastaba

- Paskaitos yra ruošiamos "as we go"
- Todėl galimos klaidos pirminėse skaidrėse
- Esminės klaidos bus pataisytos tą pačią dieną (pvz. neveikiantis kodas)
- Mažiau svarbios klaidos tikėtina irgi, visgi "final version" tą dieną, kai pradedama nauja tema
- Atsižvelgiant į "as we go", studentai kviečiami teikti savo siūlymus, pastabas ir tiesiogiai daryti įtaką kurso turiniui