

# DataCamp-kurssin rakentaminen: Johdatus yhteiskuntatilastotieteeseen

*Tuomo Nieminen, Joni Oksanen*

*14.01.2018*

## Sisältö

<b>Johdanto</b>	<b>1</b>
Johdatus yhteiskuntatilastotieteeseen (JYT) . . . . .	2
DataCamp oppimisalusta (DC) . . . . .	2
<b>Helsinki social statistics Datacamp-kurssi (HSS)</b>	<b>2</b>
HSS-kurssin rakenne . . . . .	2
HSS-kurssin lähdekoodi . . . . .	2
HSS-kurssin muokkaaminen . . . . .	4
<b>DataCamp-harjoitukset</b>	<b>6</b>
DC-harjoitusten syntaksi . . . . .	6
DC-harjoitusten testifunktiot . . . . .	7
<b>DataCamp ryhmät</b>	<b>8</b>
<b>DataCamp ja Moodle</b>	<b>8</b>

## Johdanto

DataCamp (DC) on web-alusta, jonne on mahdollista rakentaa interaktiivisia R (ja Python) kursseja. DC-kurssit koostuvat web-selaimessa ratkaistavista pienistä ohjelmointitehtävistä, joiden ratkaisujen oikeellisuus arvioidaan automaattisesti ja käyttäjälle annetaan palautetta ja vinkkejä ratkaisuyrityksistä. Datacamp on myös yritys, joka tarjoaa omia ilmaisia ja maksullisia R ja Python kursseja.

Tämän dokumentin tarkoitus on kuvata Helsingin yliopiston Johdatus yhteiskuntatilastotieteeseen (JYT)-kurssilla käytetyn DC-kurssin Helsinki social statistics (HSS) rakenne, lähdekoodi ja toiminnallisuus. Dokumentin ei ole tarkoitus korvata DC:n omaa dokumentaatiota kurssien rakentamiseen, joka löytyy osoitteesta [www.datacamp.com/teach/documentation](http://www.datacamp.com/teach/documentation).

DC-kurssin rakentaminen edellyttää Git-versiohallinnan ja GitHubin tuntemista. Tässä dokumentissa oletetaan, että näiden perusteet ovat lukijalle tuttuja.

## Johdatus yhteiskuntatilastotieteeseen (JYT)

Johdatus yhteiskuntatilastotieteeseen on sovelluspainotteinen tilastotieteen johdantokurssi. Asioita lähestytään yhteiskuntatieteiden näkökulmasta, mutta kurssi sopii muidenkin aineiden opiskelijoille.

Kurssi on jaettu kahteen erikseen suoritettavaan 5 op laajuiseen osaan, mutta muodostaa kokonaisuuden, joka käsittää yhteiskuntatieteellisen tilastollisen tutkimuksen perusteet. Osan 1 pääteema on tilastollinen lukutaito ja osa 2 painottuu tilastolliseen päättelyyn. Kurssilla on vakiintunut asema Helsingin yliopiston valtiotieteellisen tiedekunnan yhteisenä, opintojen alkuvaiheen menetelmäkurssina.

Vuonna 2015 kurssista tehtiin Suomen ensimmäinen tilastotieteen MOOC-kurssi. Sille voi osallistua kuka tahansa tilastotieteestä kiinnostunut ja sitä työssään, opinnoissaan tai vapaa-ajallaan tarvitseva. Syksyisin kurssi suoritetaan tekemällä vuorovaikutteisia, automaattisesti tarkastuvia ja palautetta antavia tehtäviä. Tarjolla on myös oppimista tukevaa lähiopetusta ja neuvontaa isoissa työpajoissa, joita vetää tiimi opiskelijaohjaajia opettajan johdolla. Keväisin ja kesäisin kurssin voi suorittaa Avoimen yliopiston verkkotenteillä. Sille voi myös osallistua ilman suoritustavoitteita milloin tahansa.

## DataCamp oppimisalusta (DC)

DataCamp (DC) on web-alusta, jonne on mahdollista rakentaa interaktiivisia R ja Python kursseja. DataCamp löytyy osoitteesta [www.datacamp.com](http://www.datacamp.com). Valmiit DC-kurssit koostuvat web-selaimessa ratkaistavista pienistä ohjelmointitehtävistä, joiden ratkaisujen oikeellisuus arvioidaan automaattisesti. Lisäksi käyttäjälle annetaan palautetta ja vinkkejä ratkaisuyrityksistä.

Kaikille avoimia DC-kursseja voi perustaa kuka tahansa. Sen voi liittää osaksi yliopistokurssia, kunhan kyseiselle yliopistokurssille voi osallistua ilmaiseksi

## Helsinki social statistics Datacamp-kurssi (HSS)

JYT-kurssia varten on olemassa oma DC-kurssi, nimeltään Helsinki Social Statistics (HSS). Se sijaitsee osoitteessa [www.datacamp.com/courses/helsinki-social-statistics](http://www.datacamp.com/courses/helsinki-social-statistics). Kurssi on avoin kaikille, joilla on DC-tunnukset.

### HSS-kurssin rakenne

DC-kurssit koostuvat kappaleista (chapter), jotka koostuvat harjoituksista. Harjoituksia voi olla kolmenlaisia: (1) koodaus, (2) monivalinta, (3) video. Melkein kaikki HSS-kurssin harjoitukset ovat koodausharjoituksia.


Koodausharjoitus koostuu neljästä osasta: (1) kuvaus/taustatieto, (2) ohjeet, (3) editori, (4) console. Koodausharjoituksen editori ja console vastaavat normaalin R editorin, kuten RStudio, editoria ja konsolia. Editorissa on yleensä valmiina koodiesimerkkejä ja harjoituksessa koodia tulee lisätä/täydentää.

### HSS-kurssin lähdekoodi

DC-kurssit ovat aina suorassa yhteydessä GitHub-repositoryyn, joka sisältää kurssin verkkosivun rakentamiseen tarvittavan koodin. GitHubin puolella DC-kurssi koostuu kappaleista (chapter), jotka ovat RMarkdown-tiedostoja. Jokaisella DC-kurssilla on lisäksi vähintään `course.yml` tiedosto, joka sisältää kurssisivulla näkyvän kurssikuvauksen ja tiedot kurssin opettajista. Ainoastaan tietyn nimiset tiedostot vaikuttavat kurssin rakentumiseen: repository voi sisältää myös tiedostoja, jotka eivät vaikuta kurssin DataCamp-ilmentymään.

FREE COURSE

Helsinki Social Statistics



Continue Course

12 hours | 0 Videos | 66 Exercises | 603 Participants | 6400 XP

## Course Description

This DataCamp course has been developed by **Tuomo Nieminen** and **Emma Kämäräinen**, under the supervision of adj. prof. **Kimmo Vehkalahti**. This course works as the Data Science module for the Social Statistics MOOC at the University of Helsinki, Finland.



**Kimmo Vehkalahti**  
Instructor at DataCamp

(Super) Social Statistician,  
D.Soc.Sci, Fellow of the  
Teachers' Academy of Uni

### 1 R and statistics

29%

Basics of R, the amazing statistical programming language. Do not be afraid of the art of programming!

Kuva 1: Helsinki social statistics DC-kurssin etusivu

1

R and statistics

29%

Basics of R, the amazing statistical programming language. Do not be afraid of the art of programming!

What is R?

✓ 50 xp

Basic tools

✓ 100 xp

Arithmetics

100 xp

Objects

100 xp

Functions

100 xp

Good arguments

100 xp

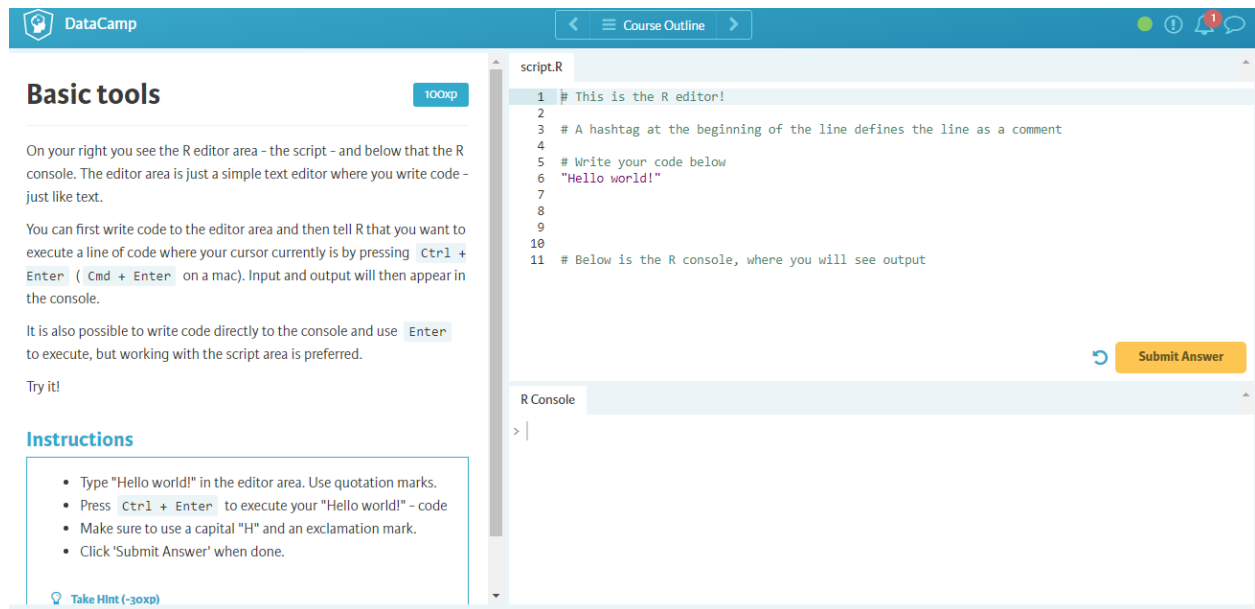
students2014

100 xp

HIDE CHAPTER DETAILS

Continue Chapter

Kuva 2: DC-kurssi koostuu kappaleista (chapter), jotka koostuvat lyhyistä harjoituksista. Harjoitusten onnistuneista suorituksista saa suoritus pisteitä (xp)



Kuva 3: Esimerkki DC-kurssin harjoituksesta: Helsinki social statistics -kurssin toinen harjoitus. Harjoitus koostuu neljästä osasta: (1) kuvaus/taustatieto (ylävasen), (2) ohjeet (alavasen), (3) editori (yläoikea), (4) console (alaoikea)

Uuden DC-kurssin (GitHub repositorio ja siihen liittyvä kurssin verkkosivu) voi luoda helposti muutamalla klikkauksella osoitteessa [datacamp.com/teach](https://datacamp.com/teach). HSS-kurssi elää osoitteessa [github.com/TuomoNieminen/Helsinki-Social-Statistics](https://github.com/TuomoNieminen/Helsinki-Social-Statistics).

## HSS-kurssin muokkaaminen

Muutokset DC-kurssin lähdekoodiin GitHubissa vaikuttavat heti itse kurssisivuun [datacamp.com](https://datacamp.com):ssa. Jokainen muutos triggeroi automaattisen kurssin rakentumisen DC:n puolella. Rakentumista ja kurssin eri versioita voi seurata hallintapaneelin avulla.

### Muokkausoikeudet

Päästäksesi muokkaamaan HSS-kurssia tulee sinulla olla mahdollisuus muokata HSS GitHub repositoryn sisältöä esimerkiksi collaborator -statuksen kautta. Lisäksi DC-tiliisi liittyvän sähköpostiosoitteen tulee olla listattuna `course.yml` -tiedostossa.

### Hallintapaneeli

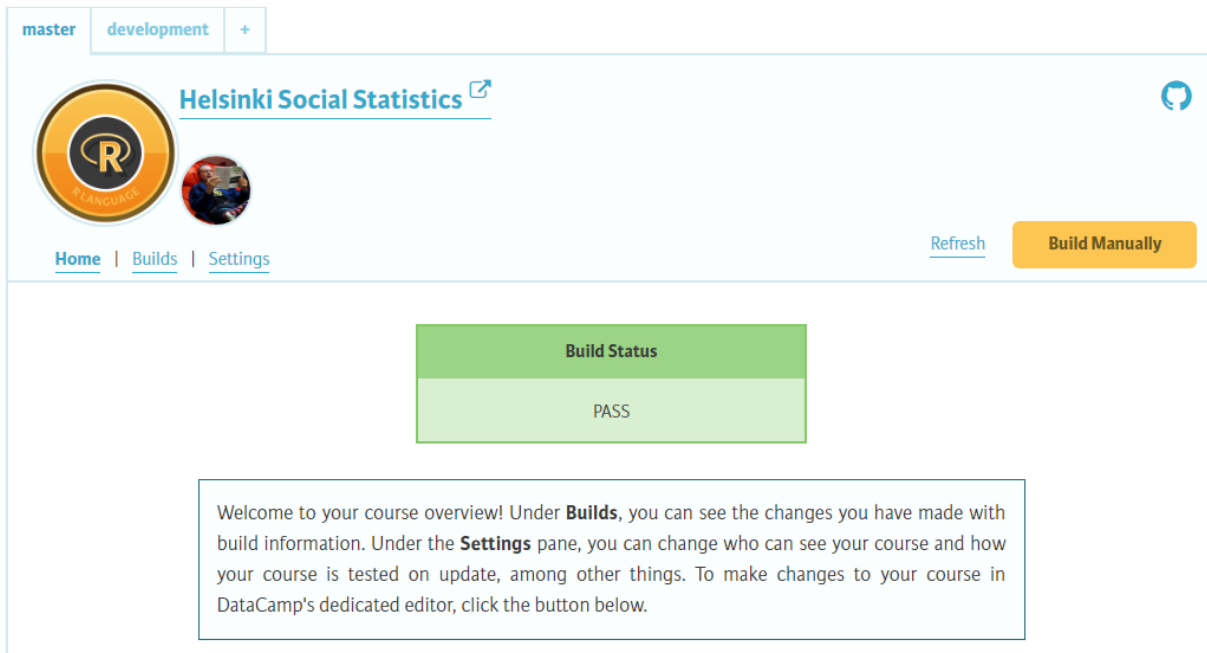
Kurssin rakentamiseen liittyvää metatietoa voi ja kannattaa seurata tähän tarkoitettulla sivulla [datacampissa](https://datacamp.com), jota kutsutaan hallintapaneeliksi. Rakennuksessa voi esimerkiksi esiintyä ongelmia, mikäli kurssikoodissa on virheitä. Hallintapaneelin kautta kurssin rakentamisen voi myös aloittaa manuaalisesti, mikä automaattisessa rakentamisessa on ongelmia. Hallintapaneeli löytyy esimerkiksi klikkaamalla Teach Dashboard -linkkiä kurssin GitHubin README-osiosta tai [datacamp.com/teach](https://datacamp.com/teach) kautta.

Branch: master	New pull request	Find file	Clone or download
<b>TuomoNieminen</b> committed on GitHub add learning objectives Latest commit 0568ec8 on 21 Mar			
data	add kyselylomake	a year ago	
docs	add learning objectives	6 months ago	
img	slight modifications	8 months ago	
.gitignore	Initial course creation commit DataCamp	a year ago	
README.md	update readme	6 months ago	
chapter1.Rmd	slight modifications	8 months ago	
chapter10.Rmd	slight modifications	8 months ago	
chapter2.Rmd	small changes	9 months ago	
chapter3.Rmd	small changes to age_counts (3.4) and plot (4.2) exercises	9 months ago	
chapter4.Rmd	small changes to age_counts (3.4) and plot (4.2) exercises	9 months ago	
chapter5.Rmd	small update 5.1.	11 months ago	
chapter6.Rmd	slight modifications	8 months ago	
chapter7.Rmd	try fix 7.1	10 months ago	
chapter8.Rmd	small changes to some exercises	10 months ago	
chapter9.Rmd	update "peculiar p-values"	8 months ago	
chapter_template.Rmd	added a template for chapter placeholders	a year ago	
course.yml	update course description (style)	6 months ago	

Kuva 4: DC-kurssien lähdekoodi on GitHubissa. Lähdekoodi koostuu RMarkdown (.Rmd) muotoisista kappaleista, course.yml tiedostosta ja mahdollisesti muista tiedostoista.

Branch: master	Helsinki-Social-Statistics / course.yml	Find file	Copy path
<b>TuomoNieminen</b> update course description (style) 26cb0bd on 22 Feb			
1 contributor			
10 lines (9 sloc)   545 Bytes         Raw Blame History			
<pre> 1 title      : Helsinki Social Statistics 2 instructors : 3   - kimmo.vehkalahti@helsinki.fi 4   - tuomo.a.nieminen@gmail.com 5   - emma.kamarainen@outlook.com 6 description : This DataCamp course has been developed by &lt;b&gt;Tuomo Nieminen&lt;/b&gt; and &lt;b&gt;Emma Kämäräinen&lt;/b&gt;, under the supervision of ad 7 university  : University of Helsinki 8 difficulty_level : 2 9 time_needed : 12 hours </pre>			

Kuva 5: Helsinki social statistics -kurssin course.yml tiedoston sisältö



Kuva 6: Jokaiseen DC-kurssiin liittyy hallintapaneeli [datacamp.com/teach](https://datacamp.com/teach) alla. Hallintapaneelistä löytyvät kurssin eri versiot sekä logitiedot kurssin rakennusyrityksistä. Hallintapaneelin kautta voi myös manuaalisesti käynnistää kurssin rakennuksen.

### HSS-kurssin kehitysversio

HSS-kurssin GitHub repositoryn main branchin sisältö määrittelee DataCampissa olevan julkistetun HSS-kurssin sisällön osoitteessa [www.datacamp.com/courses/helsinki-social-statistics](https://www.datacamp.com/courses/helsinki-social-statistics). Jokainen repositoryyn luotu uusi Git branch triggeroi uuden erillisen version HSS-kurssista DataCamp:iin. Kurssia voi siis helposti kehittää vaikkapa development nimisen branchin kautta, kuten on toimittu HSS-kurssin tapauksessa.

Linkki kaikkiin kurssin versioihin löytyy kurssin hallintapaneelistä. HSS-kurssin development-versio on osoitteessa [www.datacamp.com/courses/1935](https://www.datacamp.com/courses/1935). Tämän linkin kautta muutoksia kurssin kehitysversioon voi kokeilla käytännössä. Kun tarvittavat uudistukset on onnistuneesti toteutettu, voi development branchin mergata yhteen main haaran kanssa, jolloin uusi versio kurssista tulee julkisesti saatavaksi opiskelijoille.

## DataCamp-harjoitukset

DC-harjoitukset ovat lyhyitä ja keskittyvät yhteen asiaan kerrallaan. Tavoitteena on, että opiskelija selviää yhdestä harjoituksesta 5-10 minuutissa. Erityisesti HSS-kurssilla harjoitukset pyrkivät innostamaan opiskelijoita tilastollisen ohjelmoinnin pariin, joten turhauttavan hankalia harjoituksia pyritään välttämään.

DC-harjoitusten syntaksi on kuvattu alla. Jokaiseen harjoitukseen liittyy testejä, jotka kuvataan erikseen.

### DC-harjoitusten syntaksi

Kuten todettu, DC-kurssin lähdekoodi koostuu kappaleista (esim `chapter1.Rmd`). Kappaleet koostuvat harjoituksista, jotka määrittelee DC:n oma, RMarkdown-tyylinen syntaksi. Alla on esimerkki koodausharjoituksen syntaksista, joka koostuu (1) metatiedoista, (2) otsikosta, (3) yleisestä infosta, (4) ohjeista, (5) vinkeistä, (6) ennen harjoitusta suoritettavasta R-koodista, (7) esimerkkikoodista, (8) ratkaisukoodista, (9) testeistä.

```

“
— type:NormalExercise lang:r xp:100 skills:1
## Otsikko

Yleinen info

*** =instructions
- ohjeet

*** =hint
- vinkit

*** =pre_exercise_code
# tehtävän alustus

*** =sample_code
# esimerkkikoodi ('editor' ikkunan sisältö)

*** =solution
# ratkaisu

*** =sct
# testit
”

```

HSS-kurssin GitHub:sta löytyy valmis template uusille koodausharjoituksille.

## DC-harjoitusten testifunktiot

Oleellinen osa DC-harjoitusta ovat submission correctness testit (sct), joiden avulla testataan opiskelijoiden ratkaisuyrityksiä. Testien tulee “mennä läpi”, jotta harjoitus läpäistään. Testien tulisi testata harjoituksen jokaista vaihetta ja antaa informatiivisia viestejä silloin kun opiskelijan ratkaisu ei vastaa haluttua.

Testien avulla opiskelijan koodin tuottamia objekteja/funktiokutsuja voidaan verrata ratkaisukoodin vastaaviin. Testit kirjoitetaan testwhat R-paketin funktioiden avulla. Testeihin kannattaa tutustua esimerkiksi datacampin dokumentaation, HSS-kurssin koodin ja testwhat-paketin dokumentaation avulla: <https://github.com/datacamp/testwhat>

Harjoituksia kirjoittaessa on hyödyllistä miettiä miten harjoituksen testaaminen toimii. Useimpien koodausharjoitusten kohdalla kolmen yksinkertaisen testausfunktion avulla voi jo rakentaa riittävän monipuolisia harjoituksia. Kolme kenties hyödyllisintä testiä ovat:

- `test_output_contains("output")`: Testaa tuottaako opiskelijan koodi määritellyn tulosteen. Haluttu tulosta annetaan merkkijonona hipsujen sisällä, kuten siis tässä *output*.
- `test_object("object_name")`: Testaa onko opiskelijan koodi tuottanut nimisen objectin ja onko se testaushetkellä identtinen vastaavan ratkaisukoodin objektin kanssa.
- `test_function("function_name", args=c("arg1"))`: Testaa kutsuuko opiskelijan koodi funktiota nimeltä *function\_name* ja onko sille annettu argumentti nimeltä *arg1* sama kuin ratkaisukoodissa.

Jokaisella testausfunktioilla on argumentti `incorrect_msg`, johon erityisesti hankalimmissa tehtävissä voi ja kannattaa määritellä viesti, joka näytetään silloin kun testi ei mene läpi. Testifunktioilla on myös muita mahdollisia argumentteja, kts esim `?testwhat::test_function`. Testifunktiot antavat opiskelijalle default-viestejä silloinkin kun argumentit jättää määrittelemättä.

## DataCamp ryhmät

Kurssia varten perustetaan DC:n palvelimelle ryhmä, johon suoritustiedot kertyvät. Ryhmän avulla voidaan asettaa kurssille tehtäviä sekä ryhmän jäsenien edistymistä voi seurata omalla hallintapaneelilla.

Lisää DataCampin omilla sivuilla: <https://www.datacamp.com/groups/education>

## DataCamp ja Moodle

DC-kurssi integroidaan LTI-kytkennän (Learning Tools Interoperability) avulla Helsingin yliopiston MOOC-palvelimen (Moodlerooms) kurssialustaan, jolloin DC:lle kirjaudutaan automaattisesti samoilla tunnuksilla kuin MOOCille, ja DC-kurssin suoritustiedot saadaan MOOCin arviointikirjaan.