

# Näin delegoin työtehtäväni tietokoneelle. Katso lähdekoodi!

9.5.2018 – Blogi



KIRJOITTAJA

Juha Itkonen

Ekonomisti

Digitalisaatiolla tarkoitetaan tieto- ja viestintäteknologian laaja-alaisesta käyttöönottoa taloudessa ja yhteiskunnassa. Määritelmä voi jäädä hieman abstraktiksi. Digitalisaatio mullistaa työtä ja toimintatapoja, mutta mitä se tarkoittaa käytännössä? Seuraavassa on pieni esimerkki siitä, mitä se voisi tarkoittaa ekonomistin työssä.

Yksi monista ekonomisteille lankeavista työtehtävistä on seurata talouden viimeisimpiä käännteitä. Taloudesta kerätään säännöllisesti valtava määrä tilastoja, joista ekonomistit pyrkivät jalostamaan tietoa talouden ilmiöiden kehityksestä. Rutiininomainen datan käsittely edeltää seurantatyötä eli informaation tulkintaa ja johtopäätöksiä.

Yksinkertaistaen perinteistä työnkulkua voidaan kuvata näin: Aamulla klo 9.00 taloudesta julkaistaan uutta dataa. Ekonomisti lataa datan taulukkolaskentaohjelmaan ja alkaa tutkalla sitä. Hän tekee laskelmia ja etsii mielenkiintoisia havaintoja. Hän piirtää havainnoista kuvia ja kirjoittaa raportin (tai twiittaa).

Tietokoneet ovat tehostaneet ekonomistien ja muiden tietotyöläisten työtä monin tavoin, mutta työn tuottavuutta on yhä varaa parantaa. Digitalisaation myötä tietotyöläisten käyttöön on tullut runsaasti uusia välineitä, joiden avulla työprosessin osia on helppo automatisoida. Näin jää enemmän aikaa työn arvokkaimmille osille eli itse asiakysymysten pohtimiselle.

Automatisoinnin ei tarvitse rajoittua pelkkään datankäsittelyyn. Uusien välineiden avulla tietokoneille voi delegoida sellaista asiaosaamista, joka on aiemmin edellyttänyt ihmistyötä. Tekoälyn ja koneoppimisen avulla tietokoneet voivat tehdä itsenäisesti myös monenlaisia päätelmiä ja tulkintoja. Kun taloudesta julkaistaan uutta dataa, ekonomistin raportti on heti valmiina. [Suomen Pankin uusi lyhyen aikavälin ennustemalli](#) on esimerkki tällaisesta lähestymistavasta.

Nykyään tällaista automatisoitua tiedon jalostamista kutsutaan datatieteeksi (data science).<sup>[1]</sup> Harvard Business Review:ssä datatiedettä on kutsuttu vuosisadan seksikkäimmäksi työksi.<sup>[2]</sup>

1. Data scienceen on nykyään tarjolla useita koulutusohjelmia suomalaisissa korkeakouluissa.

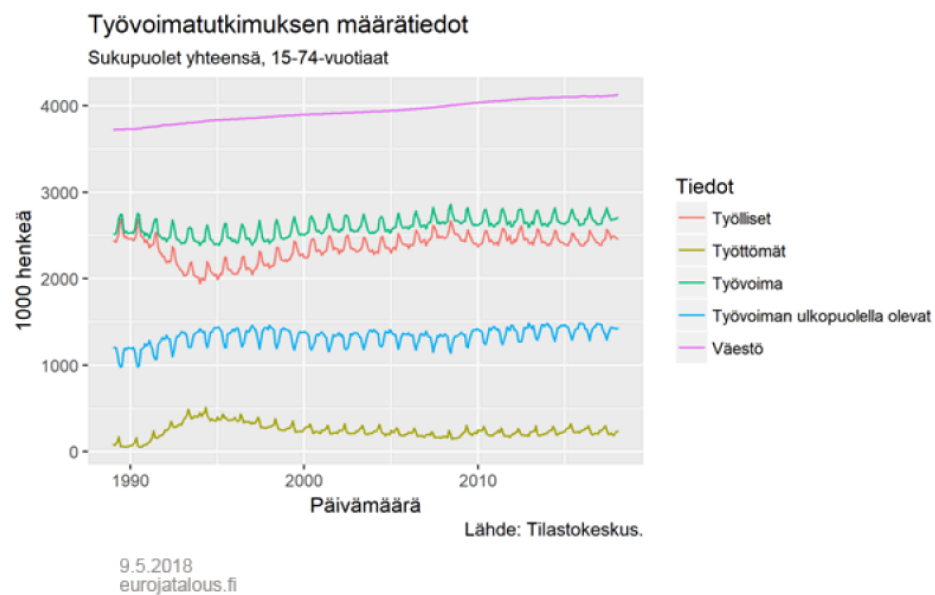
2. Ks. <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>.

Data käsittelyyn ja analysointiin löytyy monia välineitä. Avoimeen lähdekoodiin perustuvat R ja Python lienevät yleisimmin käytettyjä. Näiden ympärillä on vahva kehittäjäyhteisö, joten niihin löytyy netistä runsaasti laajennuspaketteja ja käyttöohjeita.

Automatisointia helpottaa huomattavasti myös se, että useat tilastoviranomaiset (Tilastokeskus, Eurostat ja Tulli) ovat luoneet tietokantoihinsa ohjelmallisia rajapintoja. Rajapinta on palvelu, jonka avulla tilastotiedot voi ladata suoraan ohjelmointiympäristöön yksinkertaisella tietokonekomentolla.

Seuraavassa on havainnollistettu automatisointia pienen R-ohjelman avulla. Lähdekoodi löytyy [täältä](#). Ohjelma hakee Työvoimatutkimuksen tuoreimmat tilaston suoraan Tilastokeskuksen rajapinnasta ja tekee muutamia esimerkinomaisia laskelmia sekä piirtää kuvia ja animaation.

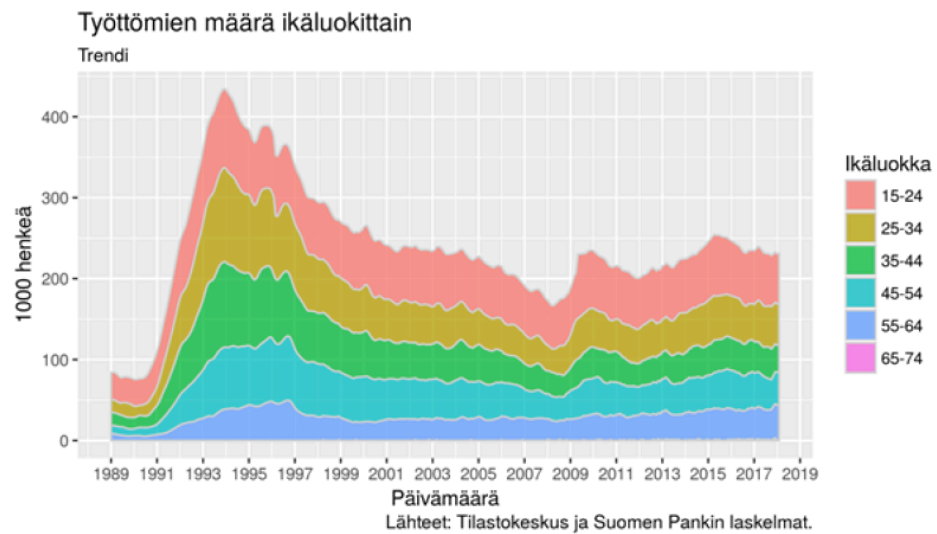
Kuvio 1.



Datan pyörittelyn ja analysoinnin koodaamisessa on monia etuja, joita ei välttämättä tule ajatelleeksi. Lähdekoodeja on muun muassa helppo jakaa, muokata ja jatkojalostaa. Laskentakaavat ovat myös kaikkien nähtävissä ja tarkistettavissa.

Lisäksi R:llä on melko vaivatonta estimoida erilaisia taloudellisia malleja tai käyttää esimerkiksi koneoppimisalgoritmeja. Kuvioista 1 havaitaan, että työmarkkinatiedoissa on voimakasta kausivaihtelua. Aikasarjojen puhdistaminen kausivaihtelusta on tilastotieteellisesti varsin monimutkainen prosessi, mutta siihen löytyy valmiita ohjelmistopaketteja R:ssä. Kuviossa 2 tilaston tietoja on esitetty puhdistettuna kausi- ja satunnaisvaihtelusta.

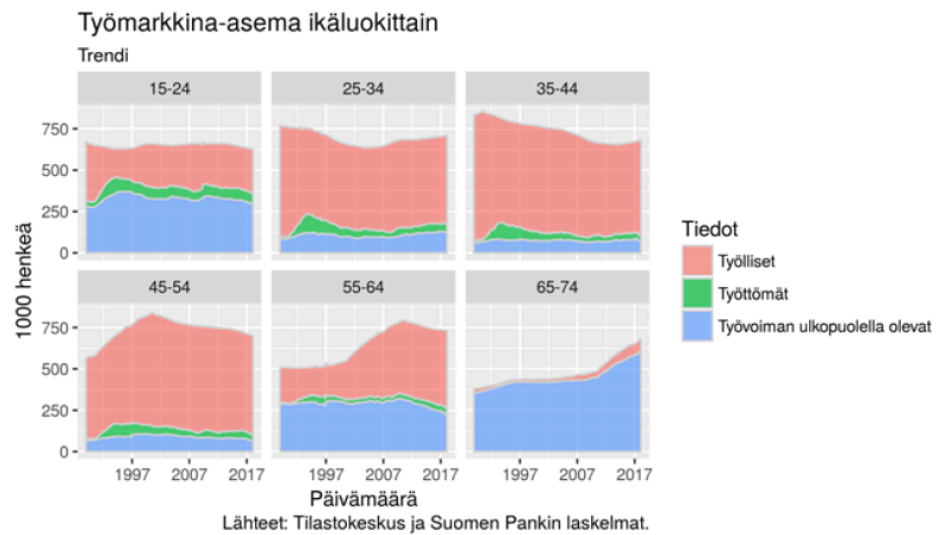
Kuvio 2.



9.5.2018  
eurojalous.fi

Tilasto-ohjelmien vahvuudet nousevat esiin moniulotteisten aineistojen tarkastelussa. Talustilastoille on ominaista, että tietoa on saatavilla luokiteltuna usean muuttujan suhteen. Esimerkiksi kuviossa 3 Työvoimatutkimuksen tietoja on jäsennetty iän, työmarkkina-aseman ja ajan suhteen. Taulukkolaskentaohjelmilla tällaisen tiedon jäsentäminen on työläästä, mutta R:ssä tilastotaulukot taipuvat vaivattomasti eri muotoon.

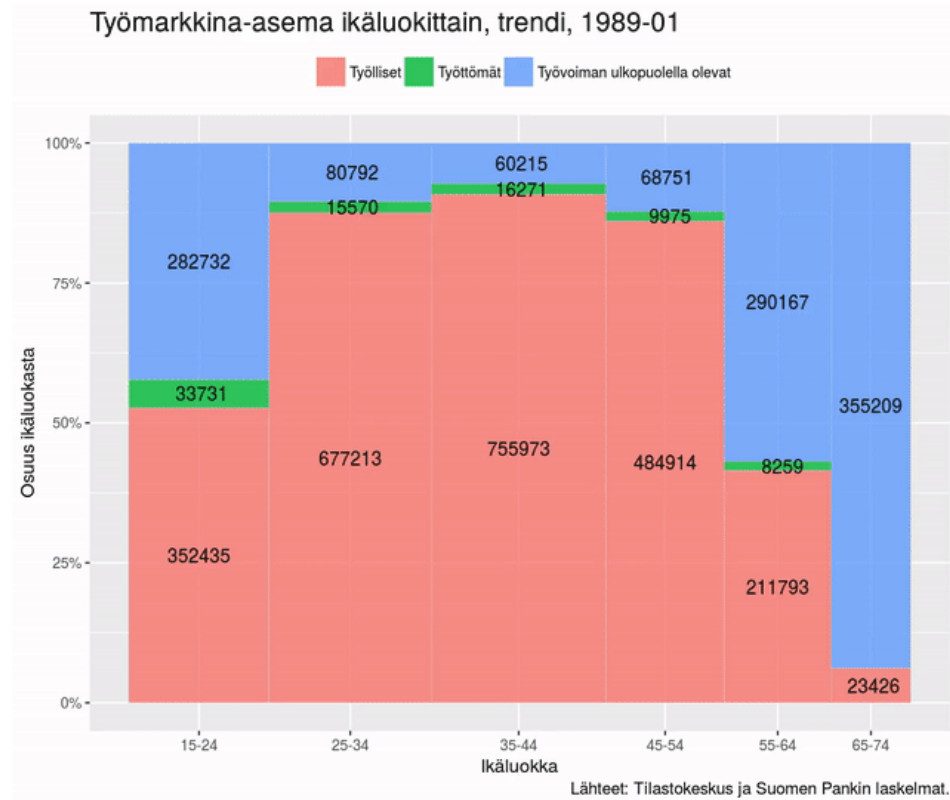
Kuvio 3.



9.5.2018  
eurojalous.fi

Koodaamalla voi tehdä asioita, joita ei muuten tulisi tehtyä. Datasta voi esimerkiksi tehdä animaatiota, jotka herättävät tilastot eloon ja saattavat tuoda uusia oivalluksia.

Alla olevassa animaatiossa suomalaisten työmarkkina-asemien kehitystä ikäluokittain on esitetty ns. mosaiikkikaaviolla. Kuviossa tiilien koot vastaavat ryhmien suhteellista kokoa ja animaatiossa juoksevat luvut kertovat ihmisten lukumäärän kussakin ryhmässä. Animaatio koostuu 350 erillisestä kuvioista, eli käsityönä animaation luominen olisi varsin työlästä.



Animaatiosta voi seurata muun muassa, miten taantumat näkyvät eri ikäluokkien työttömyydessä ja miten väestön ikääntyminen vaikuttaa talouden huoltosuhteeseen.

Digitalisaatio mahdollistaa tekemään uusia asioita ja aiempaa tehokkaammin, mutta kehitys ei tapahdu itsestään. Tietokoneet eivät vielä osaa ohjelmoida itse itseään vaan siihen tarvitaan ihmisiä, ja asiantuntijatyön automatisointi edellyttää koodaustaitoa myös asiantuntijoilta. Viime vuosina koodaamiseen on kehitetty runsaasti helppokäyttöisiä apuvälineitä ja netistä löytyy paljon ilmaisia [oppimateriaaleja](#). Oman työnsä teettäminen tietokoneilla ei ole koskaan ollut näin helppoa.

Kaikki tässä käytetyt ohjelmat ovat saatavilla ilmaiseksi ja avoimella lähdekoodilla.

## Avainsanat

[automaatio](#), [koneoppiminen](#), [tekoäly](#)