Nauka o podacima u R-u

UVOD

dr Milutin Pejović, dipl. geod. inž.
Petar Bursać, master. inž. geod.

Građevinski fakultet
2019-10-25

Dobrodošli na predmet Nauka o podacima u R-u

Anketa

- ***Da li ste ranije čuli za termin "Nauka o podacima?***
- ***Ako jeste, gde?***
- ***Zašto ste birali ovaj predmet?***

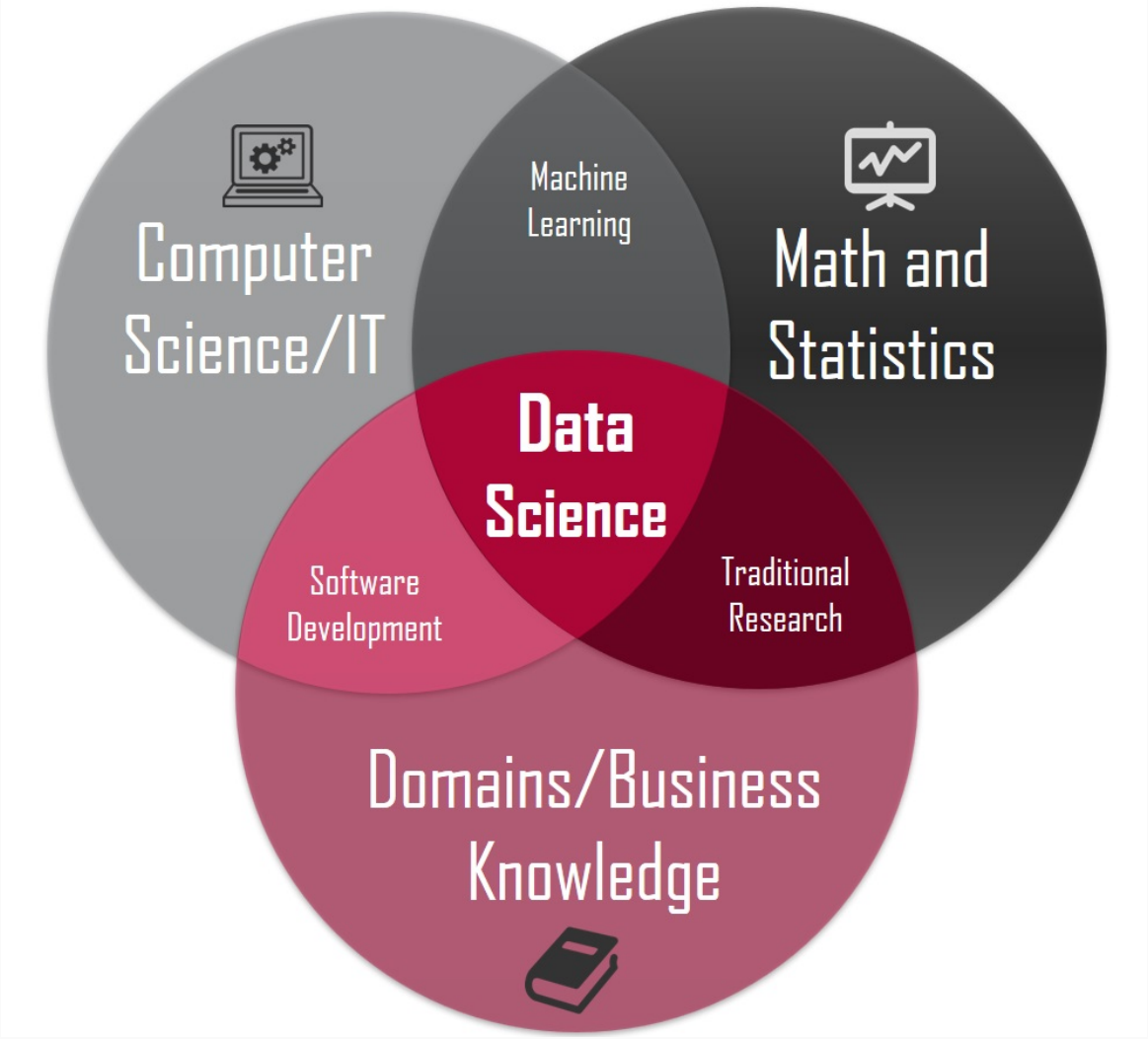
Šta je Nauka o podacima?

Nauka o podacima se bavi generičkim metodama analize podataka (vizuelizacije, transformacije podataka, statističke analize i modeliranja) kojima podaci postaju razumljiviji, a njihova informativnost veća, čime se neposredno i produbljuje znanje o pojavi koju oni opisuju.

Nauka o podacima - definicije

- *"Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data."*
- *"Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data."*
- *"Data science is the same concept as data mining and big data: "use the most powerful hardware, the most powerful programming systems, and the most efficient algorithms to solve problems"*
- *"Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data."*

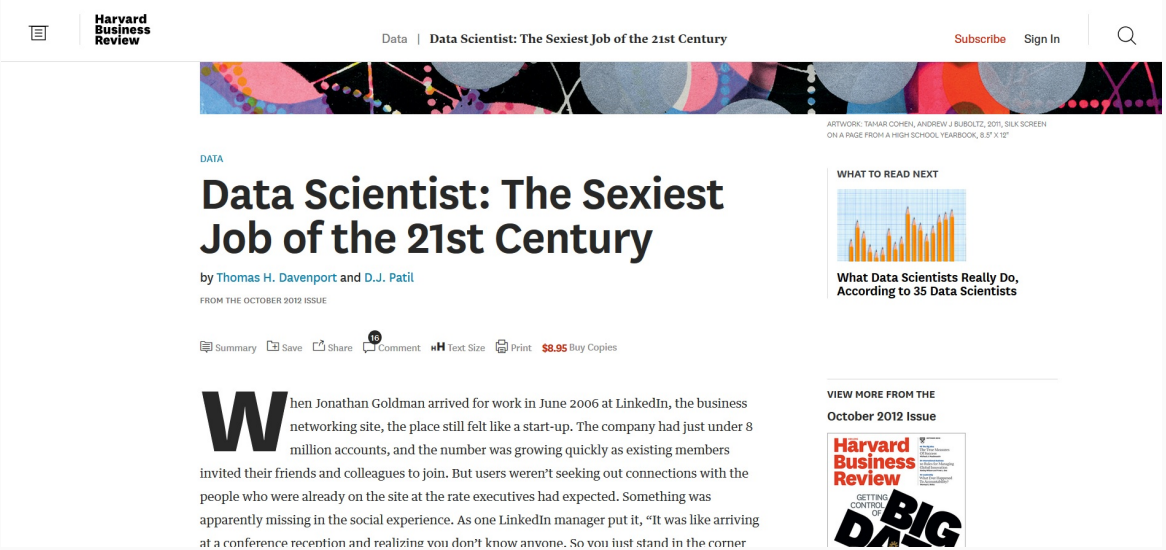
Nauka o podacima (Data Science)`



Nauka o podacima (Data Science)

- ***Relativno nova disciplina***
- ***Ima za cilj ekstrakciju znanja/informacija iz podataka***
- ***Oslanja se na statistiku, kao i na nove pristupe analizi podataka***
- ***Proistekla je iz potrebe obrade velike količine podataka (senzori, automatizovani sistemi)***
- ***Multidisciplinarna oblast***

Data Scientist



Harvard Business Review

Data | Data Scientist: The Sexiest Job of the 21st Century

Subscribe Sign In

ARTWORK: TAHAR CHEN, ANDREW J BUBOLTZ, 2011, SILK SCREEN ON A PAGE FROM A HIGH SCHOOL YEARBOOK, 8.5" X 12"

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

Summary Save Share 16 Comment Text Size Print $8.95 Buy Copies

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner

WHAT TO READ NEXT

What Data Scientists Really Do, According to 35 Data Scientists

VIEW MORE FROM THE

October 2012 Issue

Harvard Business Review

GETTING CONTROL OF BIG DA

An analytical data professional with a high degree of technical skill and knowledge, usually with expertise in programming languages such as R and Python. Data scientists help businesses collect, compile, interpret, format, model, make predictions about, and manipulate all kinds of data in all manner of ways. They're experts at both construction and deconstruction. **Even though the role of data scientist is relatively new, it's in high demand and pays well.**

Povezane oblasti ...

- ***Data mining***
- ***Statistics***
- ***Machine Learning***
- ***Big Data***
- ***Artificial Intelligence ili AI***

Spatial Data Science

- ***Bavi se problemom prostornih podataka u "Data Science" (prostorna predikcija, klasterizacija, klasifikacija, prostorna zavisnost itd)***
- ***Predstavlja logičnu nadogradnju/proširenje oblasti uzimanjem prostorne lokacije u obzir***
- ***GIS i Earth Observation***

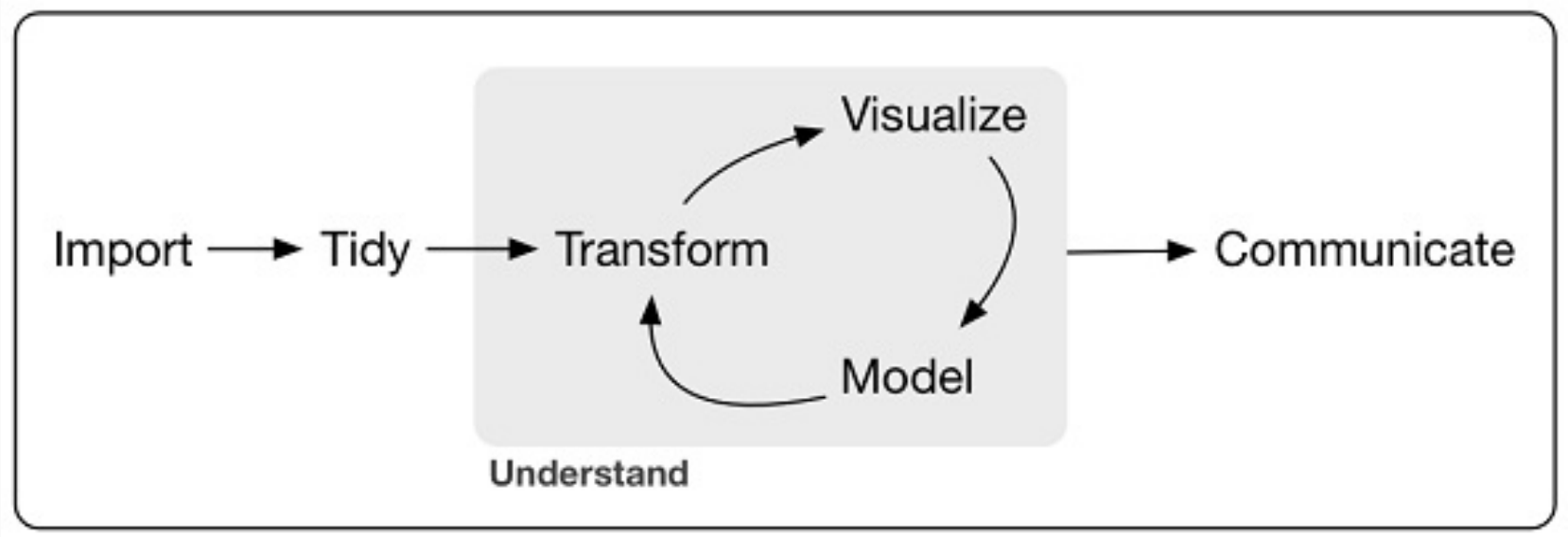
Spatial data science treats location, distance, and spatial interaction as core aspects of the data and employs specialized methods and software to store, retrieve, explore, analyze, visualize and learn from such data. In this sense, spatial data science relates to data science as spatial statistics to statistics, spatial databases to databases, and geocomputation to computation." (Luc Anselin, 2019, "Spatial Data Science" in The International Encyclopedia of Geography: People, the Earth, Environment, and Technology.

Luc Anselin, 2019
"Spatial Data Science" in The International Encyclopedia of Geography: People, the Earth, Environment, and Technology.

Nauka o podacima je proces..

"Data science is the process by which data becomes understanding, knowledge and insight"

Hadley Wickham



Šta je R?

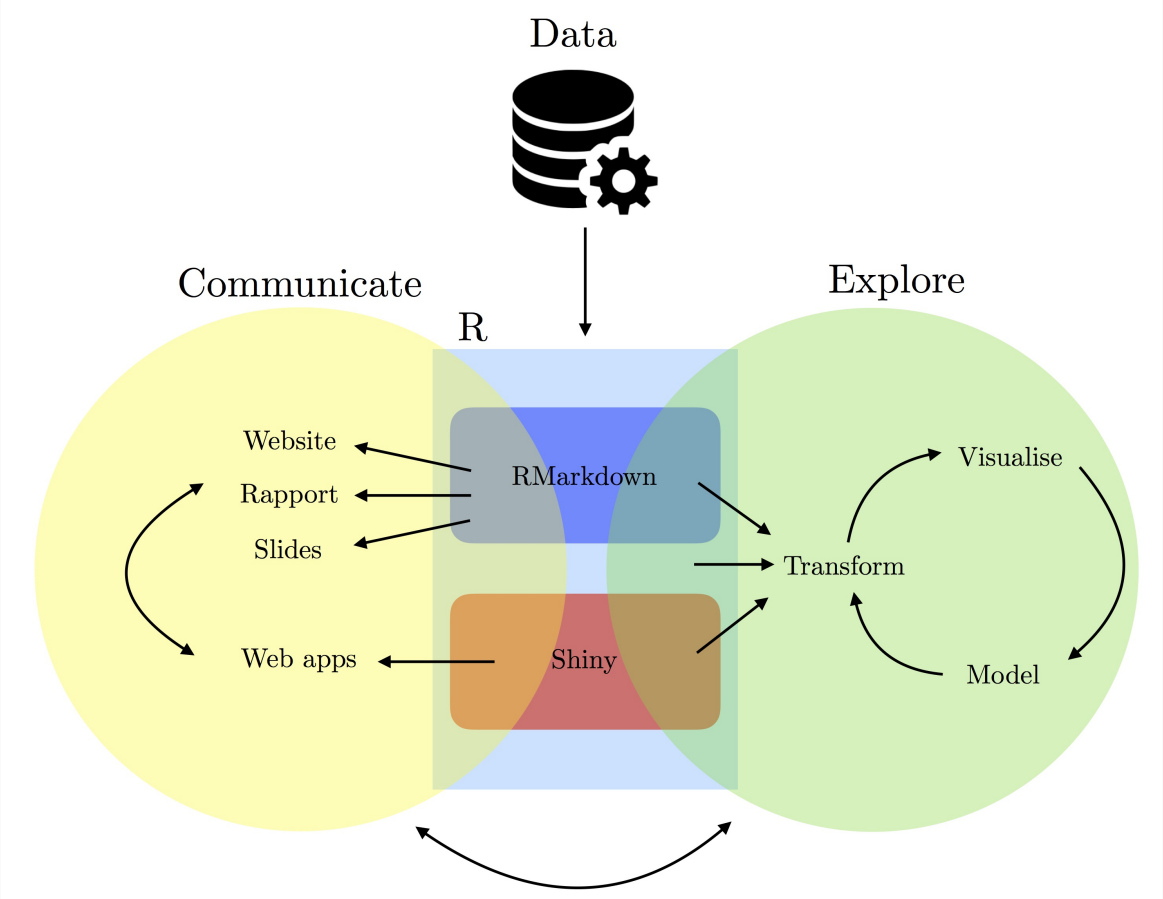
R

- ***R nije program kao Word ili Excel***
- ***R je programski jezik***
- ***Koristi se tako što se zadaju komande (koje naređuju računaru da ih interpretira)***

Zašto R?

- *"Free and Open Source" i dostupan na svim platformama*
- *Ima najbolju zajednicu korisnika (rstats on twitter, rbloggers, R Programming Language Meetup groups).*
- *Kao posledica toga, rešenja mnogih problema se brzo pronalaze na web-u (R studio, StackOverflow, R-help mailing list)*
- *Ogroman broj paketa za različite oblasti je besplatno dostupan (Cran Task View)*
- *Medju njima i fantastične alate za diseminaciju rezultata (kreiranje izveštaja u bilo kojem formatu pandoc, web i blog stranica blogdown, knjiga bookdown i na kraju paketa R packages)*
- *R studio je integralno okruženje sa fantastičnim mogućnostima*
- *R je jezik akademske zajednice i kao takav predstavlja alat za implementaciju najnovijih rešenja*
- *U okviru R-a se lako mogu koristiti drugi programski jezici*
- *R ti dozvoljava da se fokusiraš na prblem koji rešavaš.*
- *Više na Advanced R by Hadley Wickham.*

Nauka o podacima u R-u?



Book: An Introduction to Statistical Programming Methods with R

Cilj predmeta

- ***Upoznavanje sa Naukom o podacima***
- ***Upoznavanje sa R programskim jezikom***
- ***Ovladavanje savremenim metodama i alatima za rad sa podacima (manupulacija, oblikovanje, transformacija, vizualizacija, stastistčko modeliranje itd.)***
- ***Usvajanje dobre prakse (kolaborativni rada, reproduktivni izveštaji, itd.)***

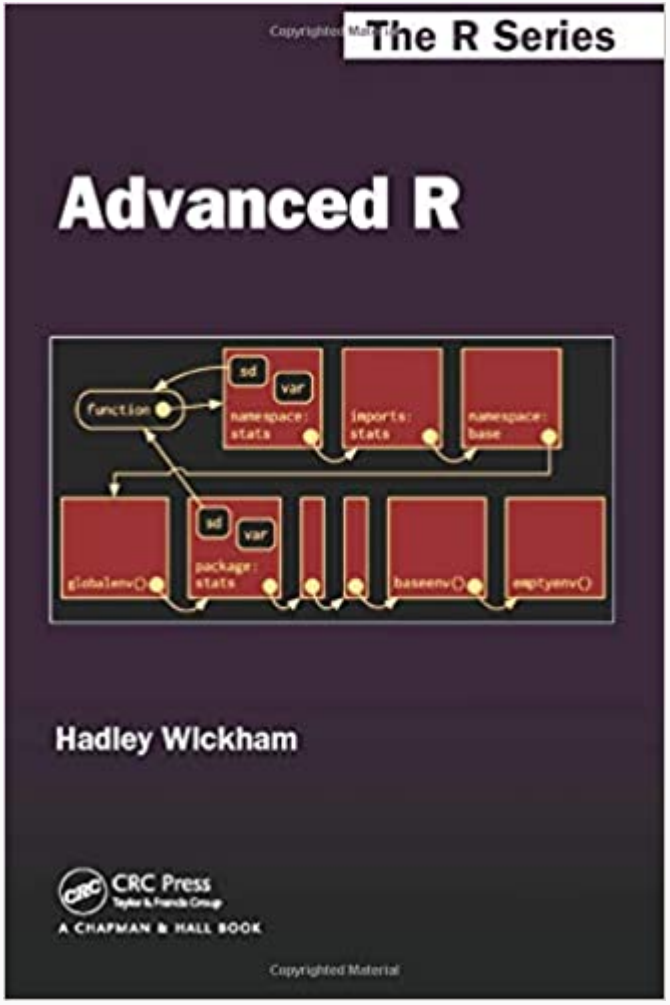
Sadržaj

- **Uvod u "R" programski jezik (R studio, CRAN, R packages)**
- **Kreiranje reproduktivinih izveštaja (R markdown)**
- **Kolaborativan rad korišćenjem GitHub-a**
- **Tipovi i strukture osnovnih podataka u "R"-u**
- **Učitavanje/ispisivanje podataka (readr, rgdal)**
- **Dodavanja, eliminacije, sortiranja, selekcije i ekstrakcije podataka**
- **Funkcionalnosti u "R"-u**
- **Oblikovanje i formatiranje podataka (tidyverse familija alata)**
- **Vizuelizacija podataka (base plot, grid, lattice, ggplot)**
- **Deskriptivna statistička analiza**
- **Prostorni podaci u "R"-u**
- **Programiranje u "R"-u (pipes, functions...)**
- **Osnove statističkog modeliranja**
- **Kreiranje "R" paketa**

Šta vas čeka

- ***Interaktivan rad***
- ***Seminarski rad***
- ***Oktobar i Novembar posvećeni učenju***
- ***Decembar i Januar posvećeni praktičnom radu***

Literatura


The R Series
Advanced R
function
sd
var
namespace:
stats
imports:
stats
namespace:
base
globalenv()
package:
stats
baseenv()
emptyenv()
Hadley Wickham
CRC Press
A CHAPMAN & HALL BOOK
Copyrighted Material

Literatura