

Statistikatarkvara – põgus ülevaade

Krista Fischer

Statistikatarkvara kategooriad

Võib jagada mitut moodi:

- Tarkvara, mis muuhulgas võimaldab ka teatud määral statistilist analüüsi teha või spetsiaalne statistikatarkvara
- Statistikakasutajale või professionaalile suunatud tarkvara
- Vabavara või kommerts
- Laiapõhjaline või spetsiifiline
-

Statistikatarkvara kategooriad

- Tarkvara, mis muuhulgas võimaldab ka teatud määral statistilist analüüsi teha
- MS Excel – nn laiapõhjaline tabelarvutustarkvara
 - Sobib hästi andmefailide haldamiseks ja esmaseks korrastamiseks
 - Lihtsad kokkuvõtted andmetest – tabelid ja diagrammid
 - Põhimõtteliselt on võimalik teha ka sisulist statistilist analüüsi
- MatLab, MathCad – matemaatikatarkvara, millel on ka statistika funktsioonid

MS Excel – mida statistik sellega peale saab hakata?

- Andmetabeli loomine (väikeste andmemahtude korral)
- Andmetabeli korrastamine:
 - Tunnuste nimed?
 - Vaatluste järjekord, suurimad-väikseimad väärtused (sorteerimine)
 - Eemalda üleliigne (read, veerud, mida pole vaja)
 - Filtreerimine – saab kiirelt uurida teatud kategooriaid objektidest

PivotTable/PivotChart

Spetsiaalne statistikatarkvara

- Tarkvara nn statistikakasutajale – inimesele, kes aeg-ajalt peab oma andmeid analüüsima, aga see ei ole tema põhitegevus ega eriala
 - Kasutajasõbralik „point-and-click“
 - Ei vaja spetsiaalset õppimist
 - Tihti vabavara
 - Tihti mõeldud teatud valdkonna-spetsiifilisemate ülesannete jaoks
 - Ei ole enamasti laiendatav ega kohandatav kasutaja spetsiifiliste vajaduste jaoks
 - Tihti n-ö „musta kasti“ printsiip – ei saa aru, mida ta täpselt teeb ja kas teeb õigesti

Spetsiaalne statistikatarkvara

- Tarkvara nn statistikakasutajale

STATGRAPHICS

STATXACT

STATISTICA

SPSS (sotsioloogidele, siiski võib pidada ka professionaalseks tarkvaraks)

Tasuta tarkvara:

<https://www.predictiveanalyticstoday.com/top-free-statistical-software/>

StatXact (CytelStudio)

Cytel Studio - [Workbook2]

File Edit View DataEditor Basic Statistics Tools Plots Nonparametrics Power&SampleSize Options Window Help

v1.csv Tabulate

Log
Workbook1
CaseData
Workbook2
v1.csv
Tabulate

pikkus:25 Value: 159

	sugu	vanus	pikkus	kaal	kohv	maiust	var	var
1	2	28	163	59	1 2			
2	1	34	179	105	1 2			
3	2	25	176	76	1 3			
4	2	24	169	49	0 2			
5	2	21	174	87	0 2			
6	2	20	178	86	0 3			
7	2	37	160	78	2 4			
8	1	23	180	76	1 1			
9	2	26	170	75	1 2			
10	2	20	171	50	0 3			
11	1	19	178	92	0 2			
12	2	22	162	49	1 3			
13	2	31	165	55	0 2			
14	1	20	179	90	1 2			
15	2	20	164	57	2 4			
16	2	37	173	81	2 3			
17	1	27	180	60	0 2			
18	2	30	157	68	6 2			
19	2	26	157	50	2 2			
20	2	18	174	80	0 2			
21	2	22	160	68	2 3			
22	2	33	160	55	0 3			
23	1	31	181	74	5 1			
24	1	28	177	75	3 1			
25	2	20	159	63	1 2			

< >

>>> [Dropdown]

Koduülesanne (statistikutele)

- Tutvuge ühe vaba statistikatarkvaraga (mis ei ole mõeldud mingiks spetsiifiliseks analüüsiks)
- Tehke üks näide mõne lihtsa andmestikuga (kas täna kasutatav „valim“, ise sisestatud fiktiivsed või päris andmed või mingid muud andmed) - arvutage ühe tunnuse keskmine ja tehke üks sagedustabel ja kui tarkvara võimaldab, siis ka üks graafik
- Tutvustage seda teistele – kas 2-3 powerpointi slaidi abil (*screenshotid*) või demonstreerige
- Võib teha 2-3 inimesega koos

Professionaalne statistikatarkvara

- Mõeldud statistikutele/andmeteadlastele
- Tihti eeldab ka mõningast programmeerimisoskust
- Vajab õppimist
- Paindlik, kohandatav, täiendatav

R, SAS, Stata, *SPSS, JMP, MATLAB*,

... Python

„Professionaalne“ statistikatarkvara – statistikutele.

Peamiselt R või SAS, viimasel ajal ka Python (mis on pigem keel, nagu ka R)
SAS on olnud pikka aega nn „tööstusstandard“ (sh ka ravimitööstuses)

<http://r4stats.com/2017/02/28/r-passes-sas/>

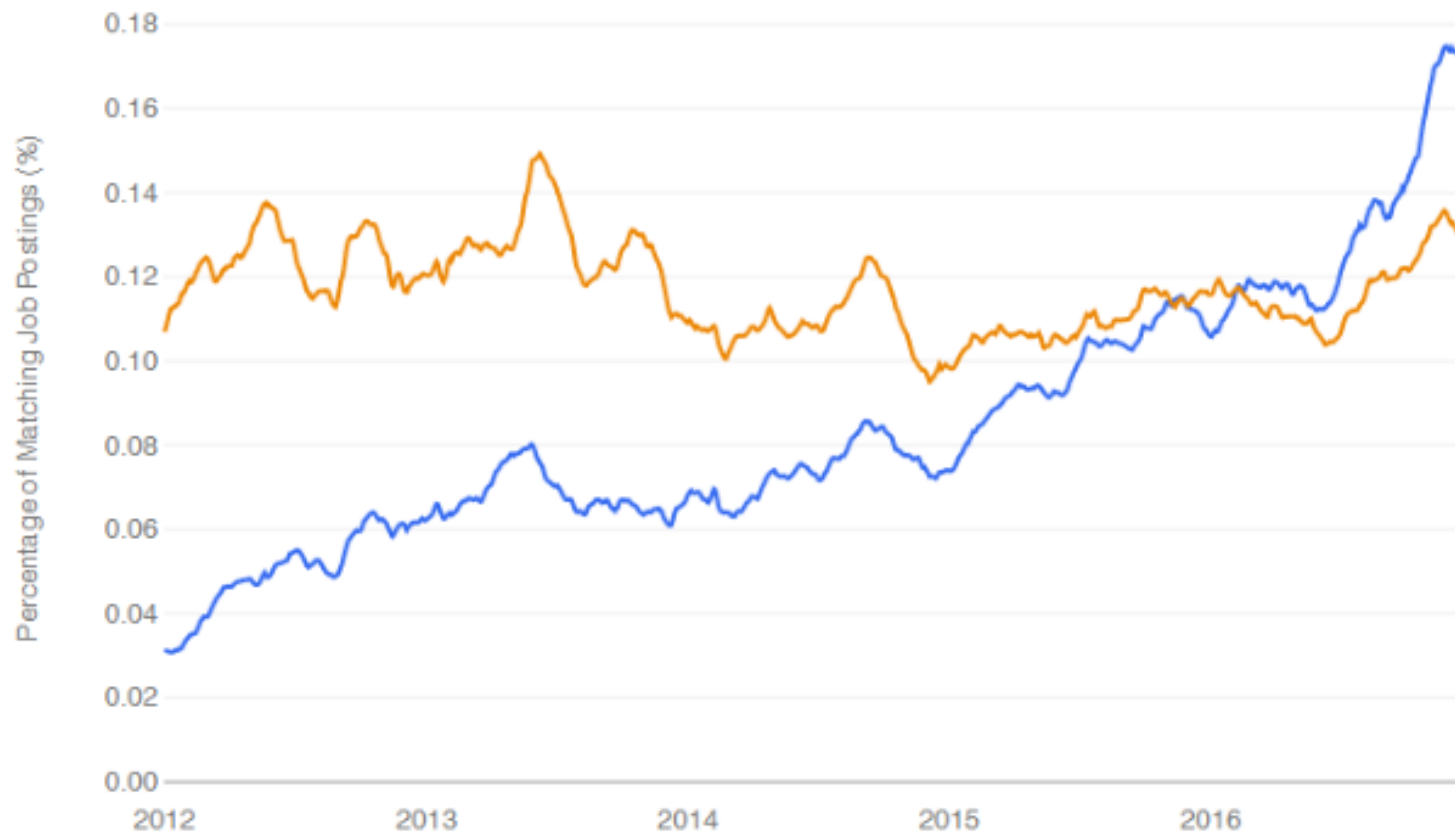
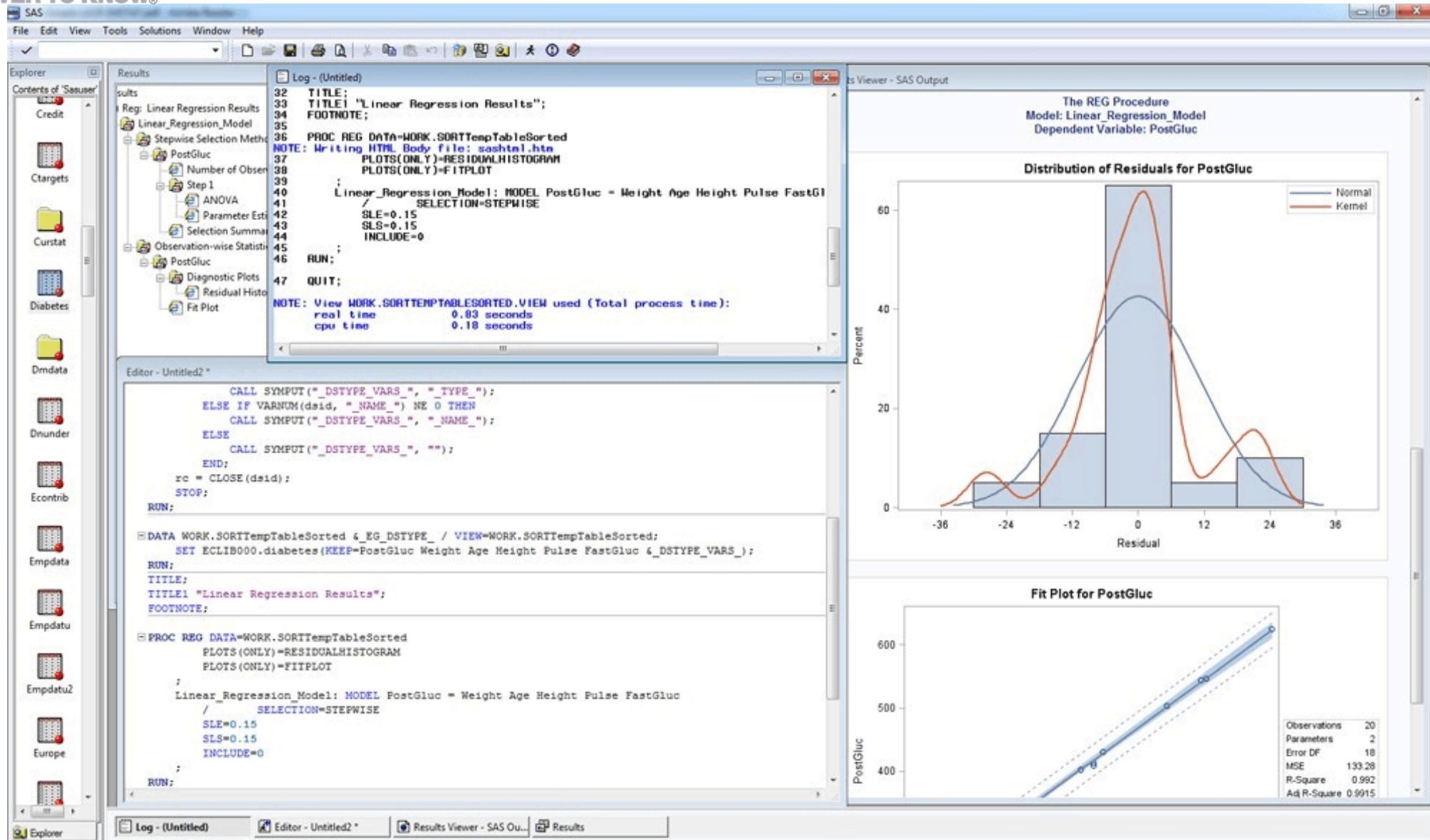


Figure 1c. Data science job trends for R (blue) and SAS (orange).



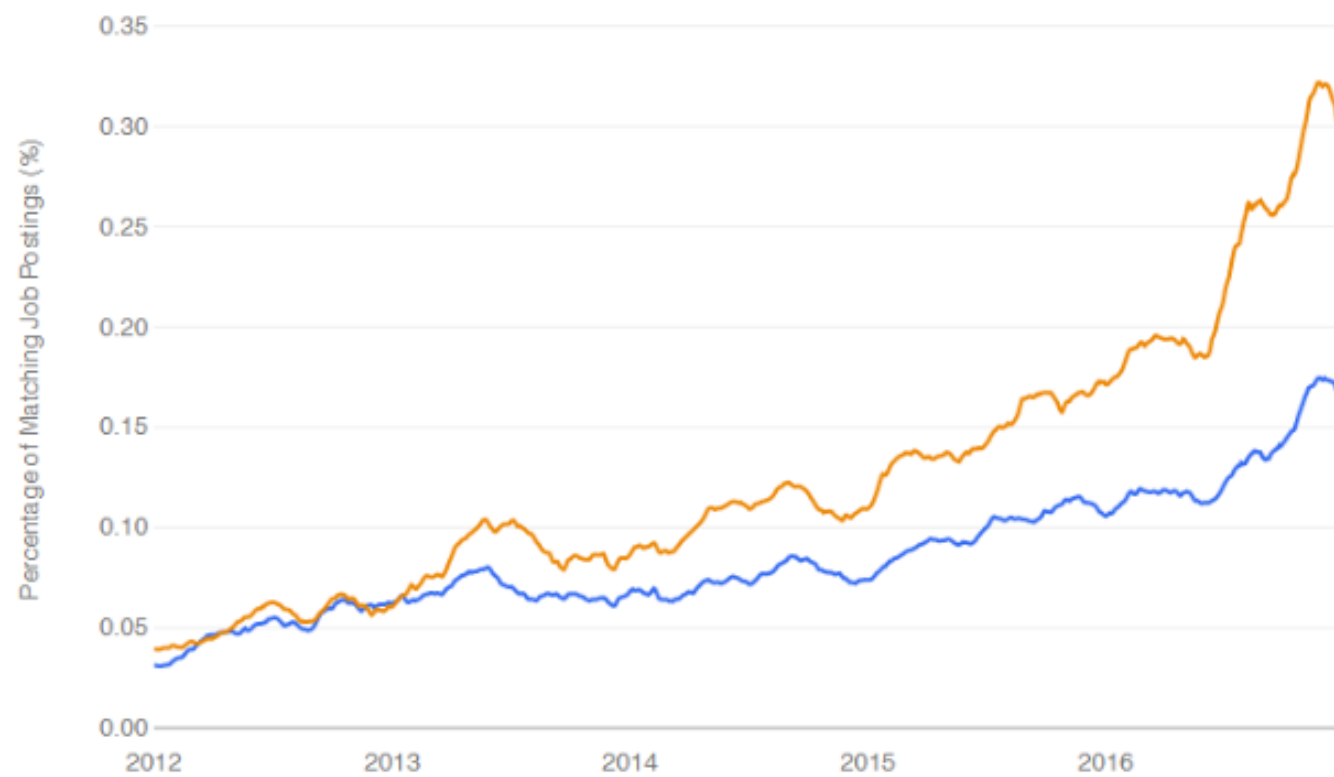


Figure 1d. Jobs trends for R (blue & lower) and Python (orange & upper).

Kokkuvõte – tarkvaranõuded statistikutele

- Professionaalse statistiku tööriist on statistikatarkvara, mis tegelikult kujutab endast pigem keelt ja keskkonda andmetöötluks
- Tänapäeval on igal statistikul vaja vähemalt mõningat programmeerimisoskust
- Samas on tihti vaja tunda ka nn tavakasutaja-tarkvara, et olla abiks muude erialade statistikakasutajatele