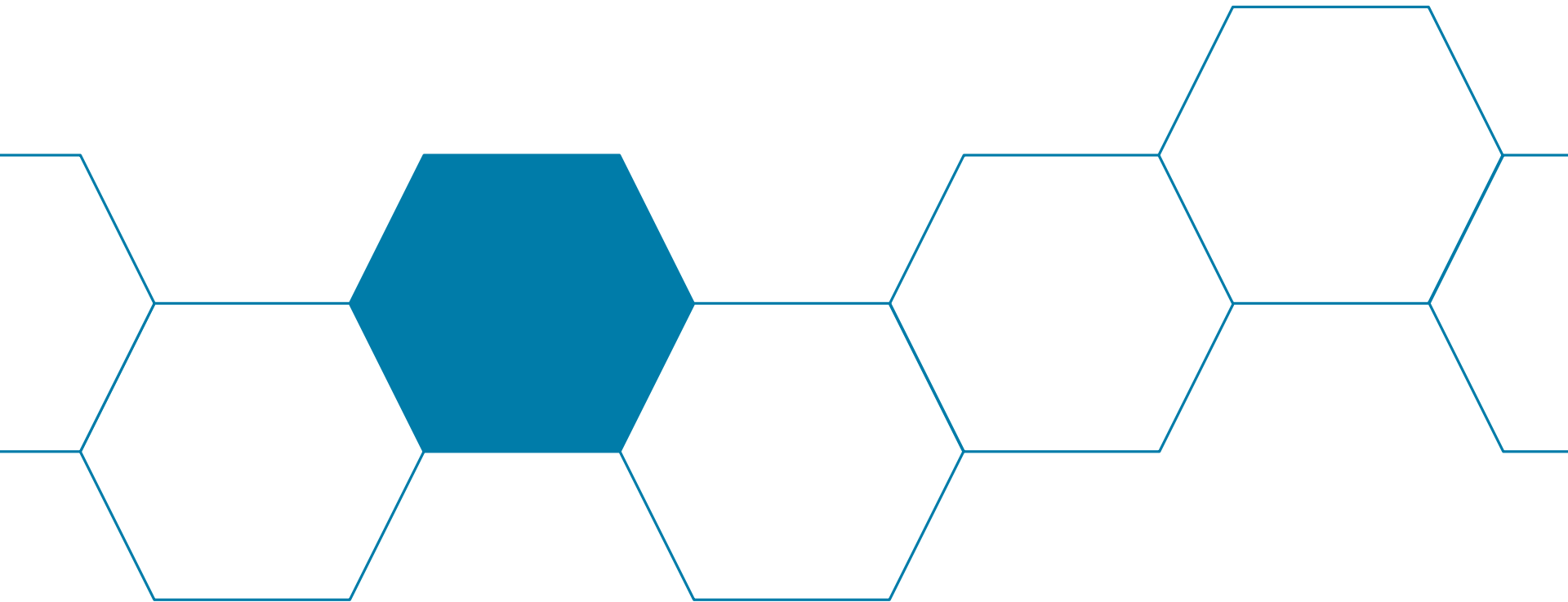


Automatiserad rapport- produktion med R



Upplägg

- Varför arbeta med reproducerbara rapporter?
- Utvecklingsmiljö och teknik
- (Teknisk demonstration)

Problemet

- Många engångsinsatser görs för att beskriva regionen och kommuner.
- Man orkar sällan uppdatera rapporter med nya data.
- Som regionbyråkrat får man ofta kommentaren – ”det är ju intressant, men hur ser det ut i min kommun?”
- Kommunerna lägger mycket pengar på konsultinköp för ganska triviala deskriptiva analyser.
- Det tar mycket tid att vara ”frågelåda” åt kommunerna (och det är ganska tråkigt).

Designval- och filosofi

- Statiska rapporter med kontextuell information.
- Undvik interaktivitet! Lägg inte analysarbetet i knäet på ovana användare.
- Erbjud kommunerna en stomme av text, diagram och tabeller som de kan vidareförädla själva.

- Se till så att både text, figurer och tabeller uppdateras utifrån data.
- Automatisera uppdateringen av databasen så mycket som möjligt.
- Gör verktyget flexibelt så att det själv går att bestämma geografiska aggregat.

Lösningen: Automatiserade kommunrapporter



- Stor efterfrågan på info om "läget" från kommunerna.
- Gjorde tidigare ett FAQ i forma av ett excelark till kommunerna, men många hade svårt att ta till sig tabelldata
- 2015 hade vi gjort ett antal moduler i R som tog fram befolkningsdiagram m.m. Idé: Kan man inte sätta ihop programsnuttarna till en rapportgenerator...
- Ägnade mycket tid under våren och sommaren 2015 år att ta fram verktyget.
- Rapporterna har legat ute på nätet sedan december 2015
- Uppskattat av kommunerna. Många frågor om de får stjäla diagram och kartor till egna rapporter.
- Rapporterna finns på:
<https://utveckling.skane.se/publikationer/rapporter-analyser-och-prognooser/kommunrapporter-for-naringsliv-och-arbetsmarknad/>
- Vi gör även kommunrapporter för Blekinge för att täcka hela NUTS2-området

Exempel:

- <https://utveckling.skane.se/publikationer/rapporter-analyser-och-prognoser/kommunrapporter-for-naringsliv-och-arbetsmarknad/>

Resultat

- Relativt högt användande bland kommunala analytiker och näringslivstjänstemän och professionella ”utvecklingsbyråkrater”. Flera fastighetsbolag har hört av sig.
- Flera regioner arbetar med att ta fram liknande rapporter. (VG, Dalarna).
- Externa önskemål om nya teman: Senast tillagda är ett bostadsavsnitt.
- Verkttyget används internt när vi snabbt behöver få ihop en kommunpresentation eller redovisa data för ett nytt aggregat av kommuner (en delregion, Sydsvenska län m.m.)
- En och annan utvecklingsbyråkrat utanför Skåne hör av sig då och då och vill ha rapporter.

Teknisk lösning:

- Lösningen bygger på programspråket R och Rmarkdown.
- Rmarkdown gör det möjligt att blanda text med kod. Figurer och tabeller skapas dynamiskt utifrån data när programmet körs.
- Rapportgeneratorn är modulärt uppbyggd: I Rmarkdowndokumentet anropas funktioner som är specialiserade på att rita diagram, kartor och skriva text.
- Uppdatering av databasen sker via skript som läser in data från SCB:s databas. (Tabeller från AF och MONA måste tyvärr uppdateras för hand)

- Data för samtliga kommuner i riket finns med för att möjliggöra jämförelser med kommunfamiljer och andra län
- Allt data lagras på kommunnivå. Andra indelningar bygger på att kommundata aggregeras
- Enkelt att skapa egna grupperingsnycklar
- Diagrammen och tabellerna i kommunrapporterna lagras även som en PowerPoint för varje kommun.
- Det går köra ut rapporterna som webbsida, MS Word eller pdf, men primärt kommunicerar webbsidan utåt.

Vad är R?



- Programspråk inriktat mot statistik
- Opensource-version av programspråket S med rötter i 70-talet.
- Modulärt uppbyggt – programmet kan expanderas med hjälp av externa bibliotek - "packages". För närvarande finns det drygt 18 000. Merparten av paketen finns i CRAN (Comprehensive R Archive Network) som står för viss kvalitetssäkring
- Stark ställning bland akademiker – nya statistikmetoder släpps ofta som tillägg till R. Även stark bland biologer och inom läkemedelsindustrin. Google och Uber är stora användare av R
- De senaste tio åren har det tillkommit kommersiella aktörer som utvecklat R för "Big Data" och skapat bättre utvecklingsmiljöer.

- Sedan några år finns det en R-version från Microsoft. R finns inbyggt i nya versioner av SQL Server och stöds av Microsofts utvecklingsverktyg.
- R och Python är de utvecklingsmiljöer som växer mest inom dataanalys. Ger större flexibilitet än till exempel SAS och SPSS.
- Kan importera data från SAS, SPSS, Excel
- Enormt utbud av visualiseringsverktyg via programpaket

Varför är det mödan värt att lära sig R (eller Python)



- Det går att få integrerade flöden i analysen. Man kan göra allt steg i analysen i samma program: Ladda hem data, datacleaning, analys, visualisering och rapportproduktion
- I och med att alla steg i analysen finns dokumenterat i kod så kan den lättare upprepas och uppdateras.

Några exempel där vi använder R:

- Framtagande av månatlig Arbetsmarknads- och konjunkturrapport.
- Kommunrapporter om näringsliv, arbetsmarknad och utbildning
- Textanalys av Arbetsförmedlingens platsannonser
- Uttag och bearbetning av data från Stativ
- Rapport över Skånes styrkeområden

Några sidor:

- <http://www.r-graph-gallery.com/>
- <http://rmarkdown.rstudio.com/gallery.html>

När använda R:

- Rapporter man kommer att upprepa
- Avancerad visualisering som till exempel Excel inte klarar – till exempel nätverk, flödesdiagram, interaktiva diagram.
- Framställning av många nästan identiska diagram.

När inte använda R

- Overkill om det är små datamaterial som ska användas en gång för att producera "standarddiagram".
- R har hyfsade GIS-funktioner, men ändå betydligt enklare med Qgis eller ArcGis...

Nackdelar med R

- Det är programmering – viss tröskel för att komma igång
- Expanderbarheten i R är en styrka, men innebär samtidigt att det finns en mängd olika sätt att skriva kod och uträtta samma sak

Resurser för att komma igång



- Ladda hem R (<https://www.r-project.org/>)
- Utvecklingsmiljön Rstudio (rstudio.com)
- Läs en introduktionsbok, t.ex. "R Programming for Data Science" av R D Peng. (leanpub.com).
- En utmärkt resurs är "R for Data Science" (<http://r4ds.had.co.nz/>). Ett "måste"!
- Tillägg för att ladda hem data från SCB: "pxweb". Data från Eurostat via "Eurostat" och från "OECD" via "oecd". Alla tre kan tankas ned från CRAN.
- Lär diagramproduktion i ggplot2 (ett diagramritningsbibliotek till R).
- Gå en nätutbildning! John Hopkins University erbjuder till exempel gratisutbildningar på coursera.com
- R-användarnätverk
- Min blogg: <https://dittochdata.netlify.com/>