



You Experienced?

Kort introduksjon til R og PowerBI



Hva er R?

- R (utt. “Ærr”) er et funksjonelt programmeringsspråk og kjøremiljø for statistiske beregninger og grafisk visualisering.
- Direkte arv fra ‘S’, utviklet av AT&T (Bell Labs) i 1975
- I dag de facto standard verktøy for modellprøving, analyse, rapportering i alle forskermiljø og blant statistikere.
- Har de siste årene fått en renessanse gjennom fokus og breddesatsing på områder som BigData, BI, og ML.
- Microsoft har gjort R til en strategisk satsing for Data Science (PBI, SQL server, BA, Cortana). Støttes også i Visual Studio.



Verktøy / ressurser

- R Server / Interpreter.
 - I utgangspunktet fritt valg; CRAN, Microsoft Server / Open
- En eller annen konsoll
 - R Studio mye brukt, IDE, funksjonsrikt, litt buggy.
 - CRAN R client (kult gufs fra 70-tallet)
 - Visual Studio
- Stack overflow (!)
- Google R Style guide



R intro

- Vektorer, datasett, utregninger, tilordninger,...
 - Alt er vektorer / arrays
 - Flerdimensjonale (matriser)
 - Alle grunnleggende operatører og funksjoner kan utføres på slike sets
 - Også støtte for strukturerte data (data frames)
 - Stappfullt av syntax quirks og særegenheter ("[...] They function in much the same way as statements in any Algol-like language.")
- Grunnleggende plot
 - Raskt, fleksibelt
 - Funksjoner og syntax farget av plotterteknologi - as in that 70's show.
- Demo

R matte

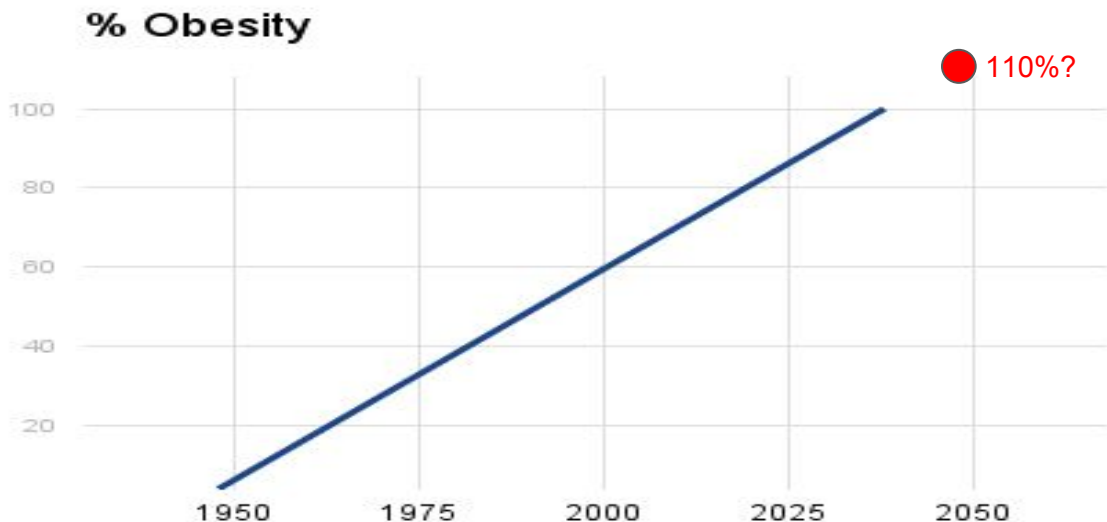


- Et vell av spesialiserte pakker, for “alt mellom himmel og jord”.
 - Generelle som prøver å avhjelpe quirks (datoformat, bedre språkkonstruksjoner; “foreach”)
 - Noen gjør veldig spesifikke utregninger, noen gjør spesifikke plots
- BigData / BI
 - Microsoft R server muliggjort stooooore datamengder, dog med lock-in som kompromiss.
- Regresjonsanalyse
 - What?
 - How?
 - When?

R Regresjonsanalyse



- Obesity journal 2008: “By 2048, all American adults will be overweight or obese.”





R matte - regresjonsmodell / analyse

```
> mfit <- lm(Co2 ~ Weight + EngDispl + Power, dieselcars)
> summary(mfit)
```

Residuals:

Min	1Q	Median	3Q	Max
-123.145	-11.743	-2.318	10.102	192.434

Avvik / Normalfordelt?

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
Weight	0.0793809	0.0005392	147.21	<2e-16 ***
EngDispl	0.0165882	0.0005645	29.38	<2e-16 ***
Power	-0.1460014	0.0061343	-23.80	<2e-16 ***

Error << Estimate?

Residual standard error: 17.96 on 23206 degrees of freedom
Multiple R-squared: 0.7068 Adjusted R-squared: 0.7068

Korrelasjon

Standardavvik /
Symmetri?



R Regresjonsanalyse

- Vær kritisk til ekstrapolering av lineære modeller
- Datagrunnlag må masseres og kvalitetssikres;
- “Outliers” må fjernes, kolonner / rader med dårlig kvalitet fjernes, kolonner med implisitt korrelasjon eller kausalitet identifiseres,...
- Det eksisterer packages for R som er mer robuste enn standard funksjoner (“rlm” / “lm”).
- Prøv ut alle idéer som kan virke fornuftig, visualiser flere fasetter.
- Test modeller - Demo!

R Plots



- Biblioteker (“packages”) for mer avansert / spesialisert plotting
- Plot med confidence
 - Eksempel fra ggplot2
- Topografi / geodata
 - Eksempel fra R-manualen
- Kommersielle packages tilbyr full 3D-funksjonalitet (plot.ly, WebGL)

PowerBI



- Produsere / publisere rapporter
 - Produsere med PBI Desktop.
 - Publisere til cloud.
 - Finpuss / manipuler på datagrunnlaget i cloud, konfigurer refresh.
- Lage rapport
 - Velg datakilde(r).
 - Finn en god måte å samle / velge felles kategoriseringer på.
 - Teste ut visualiseringer.
 - Tenk gjenbruk som “tiles” i webapper, dashboards, etc.
 - NB! Custom Visuals (<https://app.powerbi.com/visuals/>)

R / PowerBI



- R kan brukes til visualisering i PBI
 - Kan løse behov som ingen av standard visualiseringene dekker.
- R kan også brukes som datasource i PBI
 - Sømløs utvikling av BI / rapporter



You a Data Analyst?

- Komme i gang
 - Last ned R (GRATIS!)
 - Last ned Power BI Desktop (GRATIS!)
 - Aktiver PowerBI på Web / knytt opp bruker (GRATIS!)
 - Finn interessante datakilder på web (tips: www.usa.gov/statistics, www.data.norge.no)
 - Eksperimenter med R analyse og prediksjon
 - Visualiseringer i PBI (husk <https://app.powerbi.com/visuals/>)
 - Be Impressed!
 - Kildekode og datafiler på:
<https://github.com/KetilParow/R-PBI-Talk>

