



Velkommen

til STK1000, Innføring i anvendt statistikk

**Kathrine Frey Frøslie. Statistiker, PhD.
Nasjonal kompetansesenter for kvinnehelse
Oslo Centre for Biostatistics and Epidemiology**



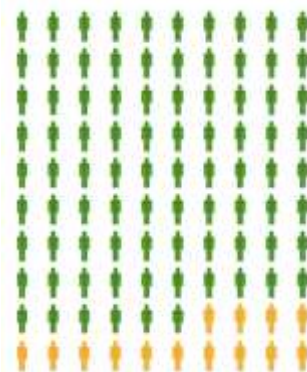
**statistikk
er historien om de mange**



Blant 100 barn
(under 5 år)
som får meslinger:



Blant 100 barn
som får MMR-vaksine:



Milde symptomer – behandles hjemme

Moderate komplikasjoner – legekonsultasjon

Alvorlige komplikasjoner – akutt legehjelp (sykehusinnleggelse)

Blant 100 barn
(under 5 år)
som får meslinger:



1 av 100
3 000 av 1 000 000
2 av 1000
1 av 100 000

Feberkrampe
Blødningssykdom
Hjernehinnebetennelse (hjerneskade/død)
Akutt, dødelig hjernehinnebetennelse

Blant 100 barn
som får MMR-vaksine:



1 av 1 000
26 av 1 000 000
Maksimalt 1
Ingen



**Vi må undersøke mange
for å trekke ut essensen:**



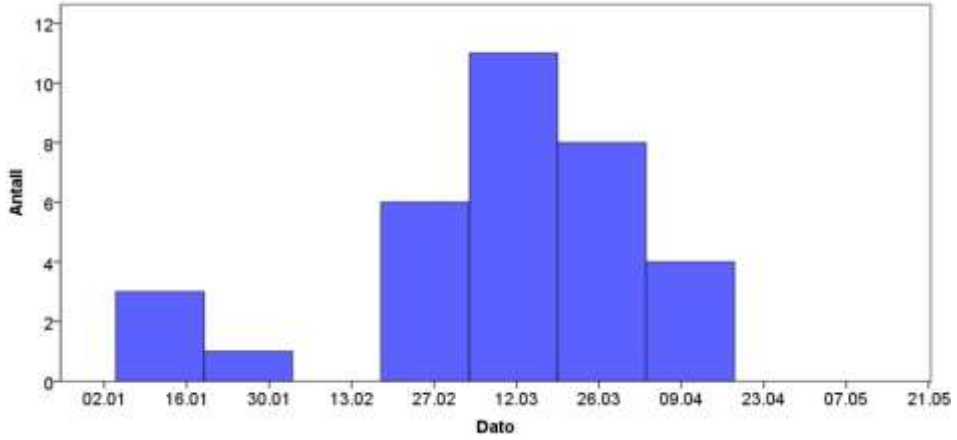


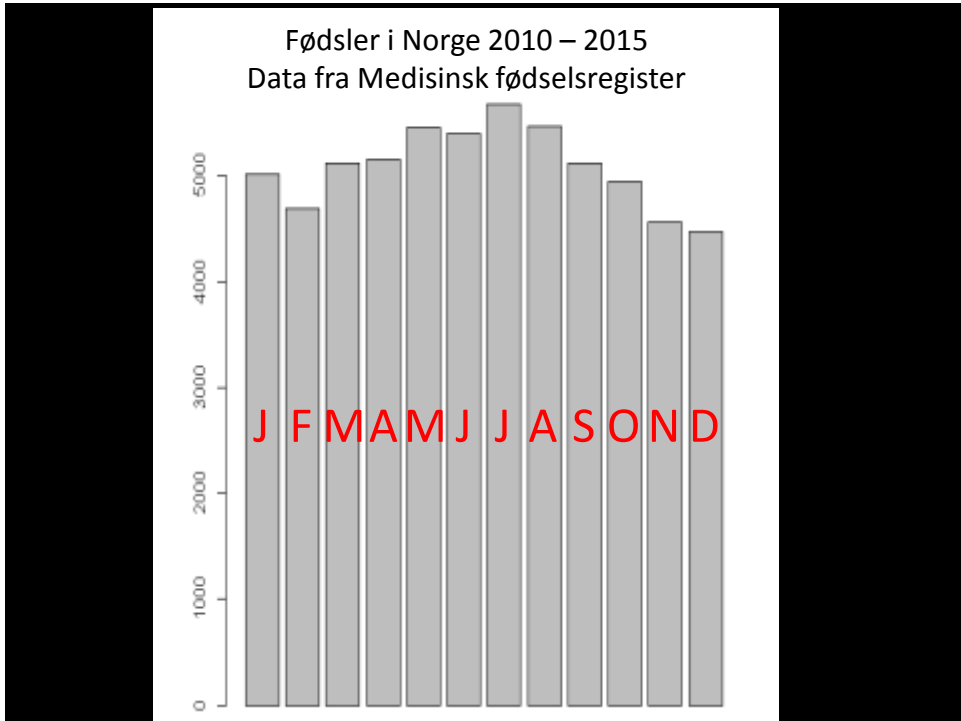


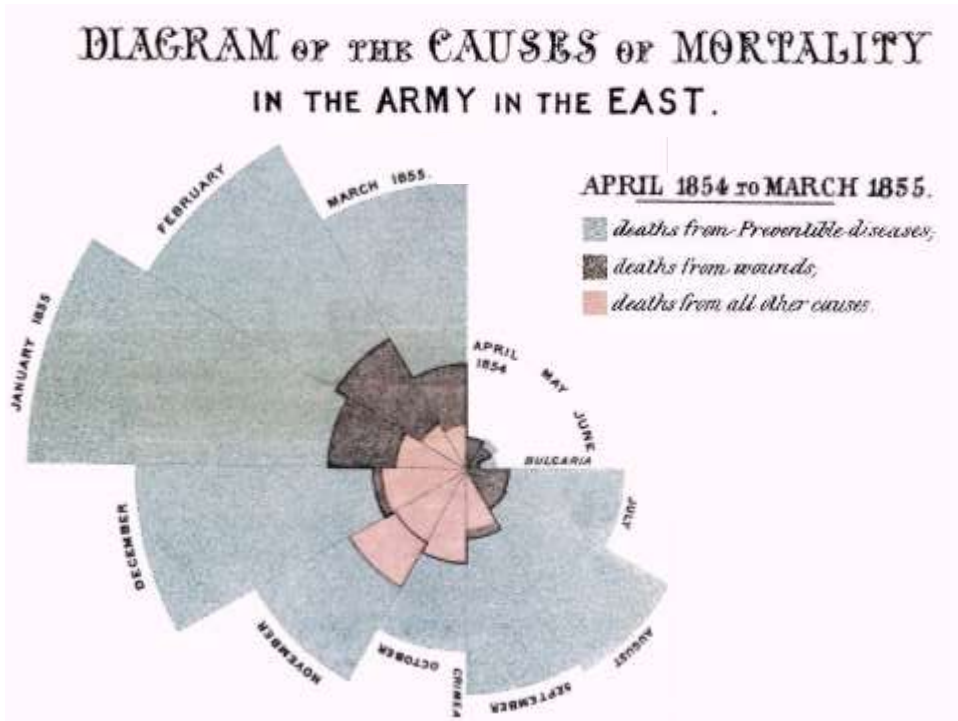
Florence Nightingale tending to wounded soldiers. National Library of Medicine

**Oppsummeringene er viktige
(og noen er vakre)**

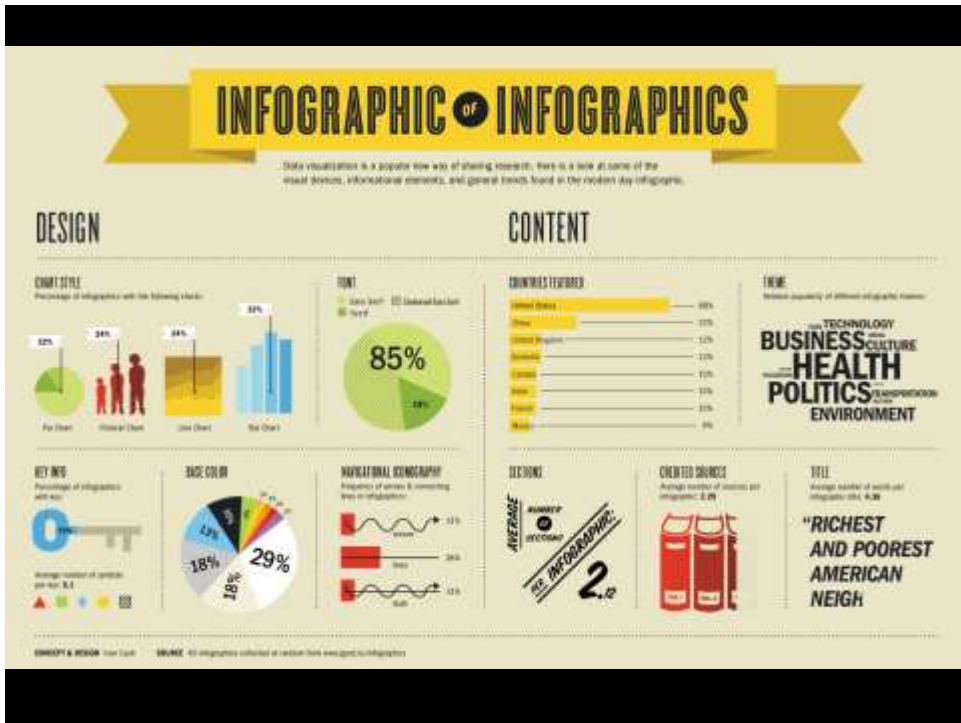
**Slike oppsummeringer er hva
folk flest forbinder med
statistikk.**







**Infografikk, InfoVis,
Information visualization:
«Information is beautiful»**



**Igangsetting av tiltak krever
ofte bruk av statistiske analyser
(både rene oppsummeringer,
men også mer
sofistikerte analyser)**



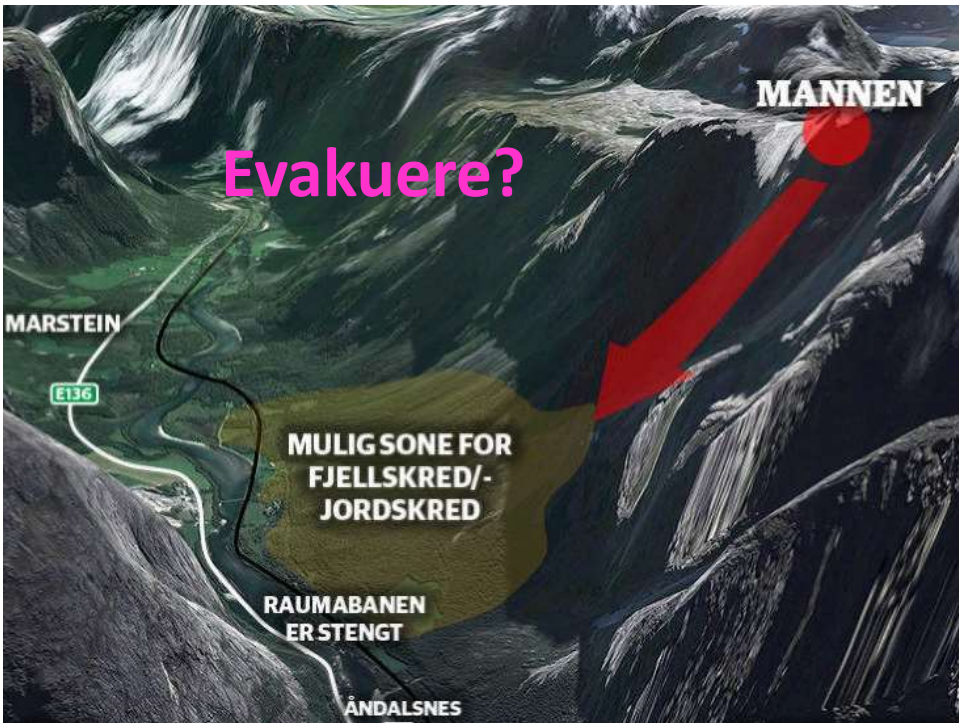
**Krybbedød: Fra 140 per år til
under 20 per år etter kampanje**



Kaste eller drikke?



Ekstremværvarsel





Big data



BIG DATA – THE PROMISE

SORTING THE REALITY FROM THE HYPE

The Hype

- Turning data flood into transformative insight
- Exploration without preconceived notions

The Reality

- Big Data application stack is complex
- Data is highly fragmented
- New skill sets
- Difficult to know where to start

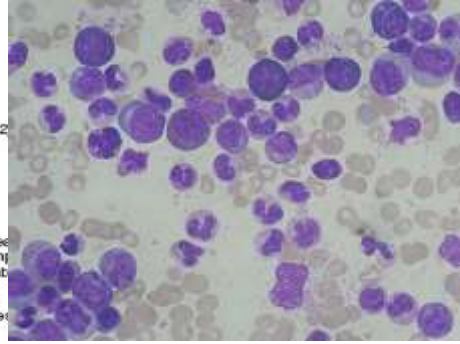
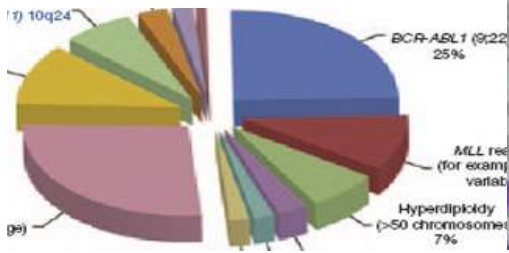
How does Big Data help my Business?
 How much will it cost?
 What's the ROI? How risky is it? When will I see results?

4

© Hitachi Data Systems 2013

Personalized medicine: Chronic myeloid leukemia (CML)

Cancer of the white blood cells. In Western countries it accounts for 15-20% of all adult leukemias



CML is now treated with targeted drugs called tyrosine kinase inhibitors (TKIs). Before TKIs, median survival time for CML patients was 3–5 years from time of diagnosis.

With the use of TKIs, the overall survival rate is about 90% after five years.

Promotion

Get to know yourself

The new web app **KG Land**, powered by **Karimnagar.com** psychology and genetics to help personal development

DNA & you

As humans, we spend our lives working and getting to know other people, but often we don't really know ourselves. The app **KG Land**, powered by Karimnagar.com, is designed to help you understand your genetics, psychology and personality, and to help you to be the best in a personal and social way.

Thanks to the app, you will be able to know your DNA and you will be able to know your personality and your social life. The app will be able to help you to be the best in a personal and social way.

As humans, we spend our lives working and getting to know other people, but often we don't really know ourselves. The app **KG Land**, powered by Karimnagar.com, is designed to help you understand your genetics, psychology and personality, and to help you to be the best in a personal and social way.

Thanks to the app, you will be able to know your DNA and you will be able to know your personality and your social life. The app will be able to help you to be the best in a personal and social way.

As humans, we spend our lives working and getting to know other people, but often we don't really know ourselves. The app **KG Land**, powered by Karimnagar.com, is designed to help you understand your genetics, psychology and personality, and to help you to be the best in a personal and social way.

Thanks to the app, you will be able to know your DNA and you will be able to know your personality and your social life. The app will be able to help you to be the best in a personal and social way.

As humans, we spend our lives working and getting to know other people, but often we don't really know ourselves. The app **KG Land**, powered by Karimnagar.com, is designed to help you understand your genetics, psychology and personality, and to help you to be the best in a personal and social way.

Thanks to the app, you will be able to know your DNA and you will be able to know your personality and your social life. The app will be able to help you to be the best in a personal and social way.

Can your genes decide if you're good at saving money?

based on their personality and their own self-perception," says Kyriakos Kokkoris, co-founder and CEO of Karmagenes.

If the users want to know even more about themselves, they can then send off a swab of saliva to the Karmagenes HQ in Switzerland for a DNA test. They'll then receive a 37-page detailed report in the form of an e-book, explaining what their DNA and environment say about their personality and behavioural characteristics.

"Having a psychological test coupled with genetic analysis through gamification makes the science more comprehensive and user friendly. It provides meaningful information based on both nature and nurture," says Kokkoris.

The team consists of five scientists, who have PhDs ranging from neuroscience and genetics to molecular biology – plus

a psychologist. They joined forces in 2013 with the initial idea to connect people through DNA and bring science to the general public. They used their collective knowledge to develop the concept.

Understanding yourself is not only beneficial on a personal level, but also on a professional one. "Genetic analysis can show us our potential and hidden talents. You might already know you're good at organisation, but you never realised you have other inherent skills such as being decisive and confident," explains Kokkoris. "If your DNA shows you have these skills, you're more likely to believe in yourself and make better professional decisions. Our goal is to promote the positive aspects of DNA."

The web app is still in the beta testing stage but, according to Kokkoris, the next step is to create a social network that is based solely on personality. Is this the future of friendship? Either way, Karmagenes helps you get to know the most important person in your life – yourself.

www.kg.land

Hjernen er en stor statistikk-kalkulator





P(Inntil døden skiller oss ad)

MatNat, laveregradskurs, 2010

A	B	C	D	Stryk
12%	23%	34%	19%	12%

STK1000 H2016?

Ambisjoner?

Antall timer totalt?

Kjønn?

Kursets rykte?

Jevnt og trutt vs

Liker du appelsiner?

skippertak?

Tidligere karakterer?

Tidsnok

til timene?

Motiverende Praktiske oppgaver?

eksempler?

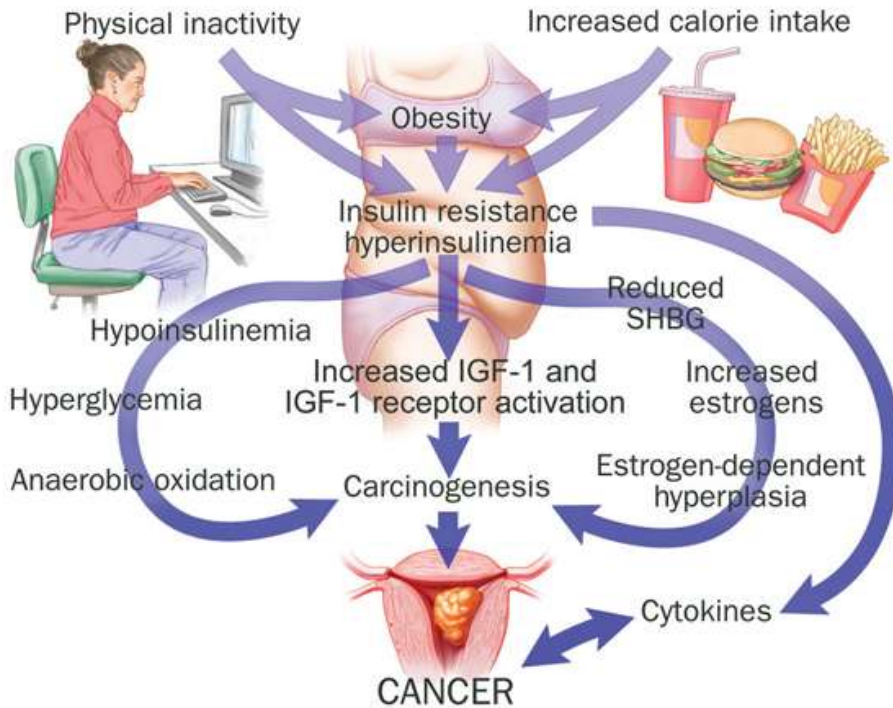
Benytte

gruppetimer?

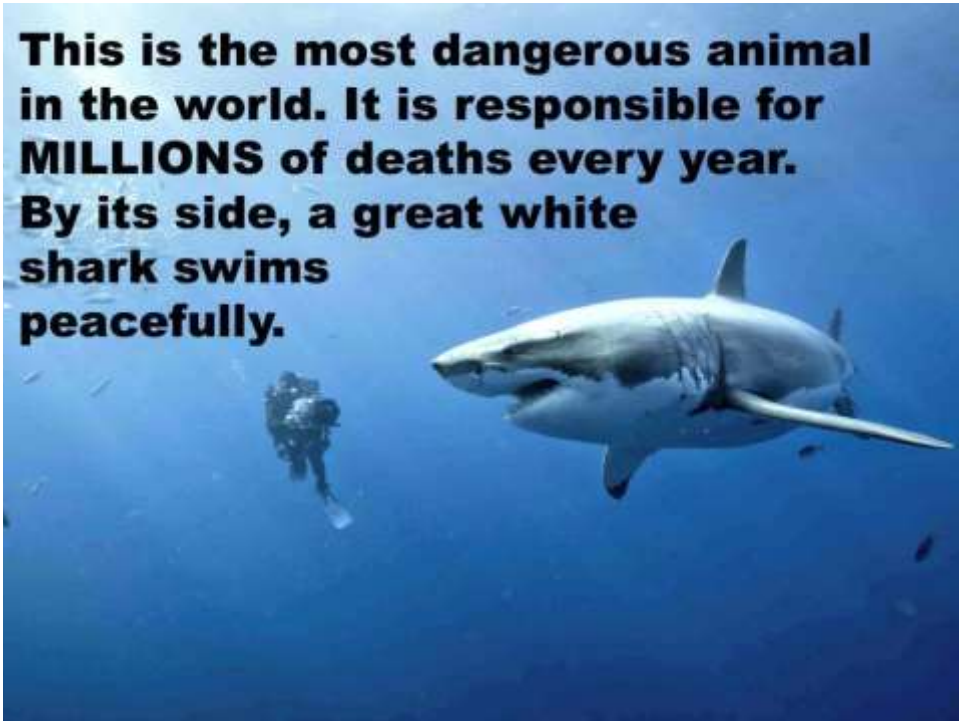
The screenshot shows a car listing website with various filters on the left and four car listings on the right. The filters include 'Merket', 'Avgiftsklasse', 'Drivstoff', 'Farge', 'Hestekrefter', 'Seter', 'Hjulsett', 'Girkasse', 'Årsmodell', and 'Kilometerstand'. The listings are as follows:

Model	Year	Price
Audi A3 Sportback 1.8 TFSI	2012	173 999,-
Audi A5 NY FWD - Audi A5 Sportback 2.0 TFSI 150 hk multitrak	2011	274 999,-
Audi A4	1996	18 438,-
Ford Focus 1.3 DSC T&E	1979	191 538,-





This is the most dangerous animal in the world. It is responsible for MILLIONS of deaths every year. By its side, a great white shark swims peacefully.



Vi gjør statistisk analyse hele tiden uten at vi er det bevisst.

Ofte er våre subjektive beregninger, det vi tror, **FEIL**.

*Statistikk brukes for å finne ut om verden er slik vi *tror* den er*

Skal vi formalisere beregningene, trenger vi notasjon, språk og verktøy.

Vi må uttrykke oss presist om statistiske analyser

Notasjon, språk og begreper

Vi må undersøke mange for å trekke ut essensen

Hvor mange er «mange»?

Oppsummeringene er viktige

Hvilke oppsummeringer?

Figurer? Oppsummeringstall?

Igangsetting av tiltak?

Hva ønsker du å oppnå?

- Hvilken respons?

- Hvilken forklaringsmodell?

Grisete datasett?

Datavasking, forberedelser til analysen

Svære datasett?

Big data? Forklare? Predikere?

Data-behandling

Programvare (R)

IRL:	Web:
Forelesninger	Kursets nettside
Grupper	Facebook-gruppe
Papir:	
Boka: Introduction to...	
Tall forteller	IT:
	R
Praktisk:	Nettskjema
Datainnsamling	Excel
Kahoot	www.tallforteller.no

STK1000: OVERSIKT

I boka

1. Planlegging og design, data-innsamling og ulike typer data	3.1-3.3, 3.5
2. Enkel oppsummering: Beskrive data (deskriptiv analyse) med figurer og oppsummeringstall	1.1-1.3
3. Beskrive enkle (bivariate) sammenhenger: Krystabeller, boxplot, scatterplot	2.1-2.2
Chi-square, t-test/W signrank test, korrelasjon	9.1,7.2/15.1,2.6 (10.1)
4. Regresjonsanalyse: Outcome, covariate, residuals To explain or to predict?	2.4, 2.5, 2.7, 10.1-10.2,
Confounder/mediator/collider, variable selection	11.1-11.2
✓ Populasjon – data – modell. Normalfordelingen CLT og N-fordeling, 5.1 Binomisk og N-fordeling, Poisson og N-fordeling	1.4 5.2
✓ Sannsynlighetsbegrepet	4.1-4.2,4.5
✓ Stokastiske variabler, sannsynlighetsfordelinger og –tettheter	(1.4, 5.1-5.2) 4.3-4.4
✓ Inferens: Konfidensintervall og Hypotesetesting (inkl p-verdi, styrke)	3.4, 6.1 6.2-6.4
Ett-utvalgs t-test	7.1
To-utvalgs t-test og Wilcoxon's rangsum-test	7.2, 15.1

STK1000: OVERSIKT

1. Planlegging og design, data-innsamling og ulike typer data

I boka

3.1-3.3, 3.5

**Første praktiske oppgave:
Appelsinskrellekonkurransen**

**Forskningsspørsmål/
Problemstilling:**

<https://nettskjema.uio.no/answer/74752.html>





**Neste forelesning,
fredag 26. august
kl 10.15-11:**

**HåvaRd Kvamme
gir en innføring i R**



<https://www.r-project.org/>