



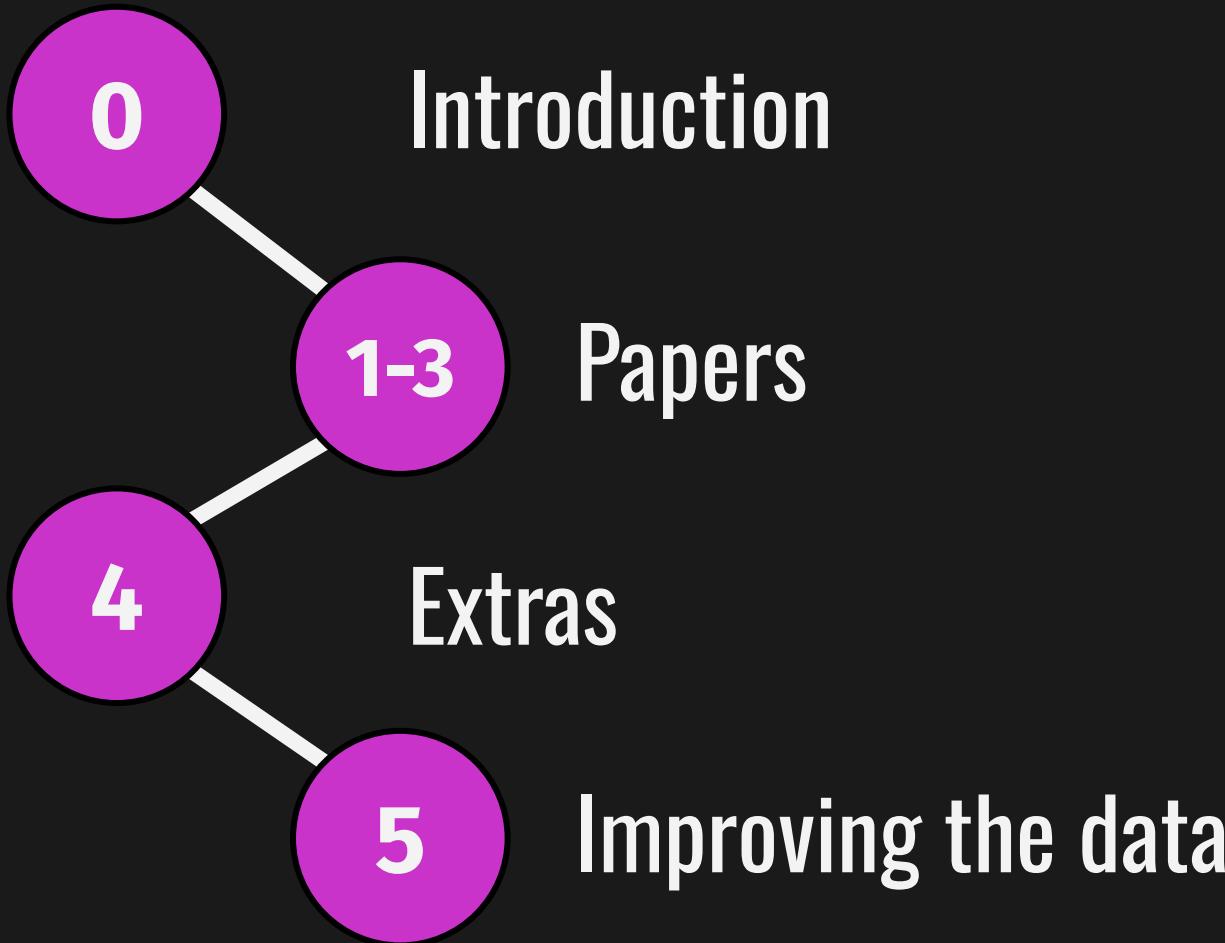
UiT The Arctic University of Norway

SOCIAL NETWORK ANALYSIS

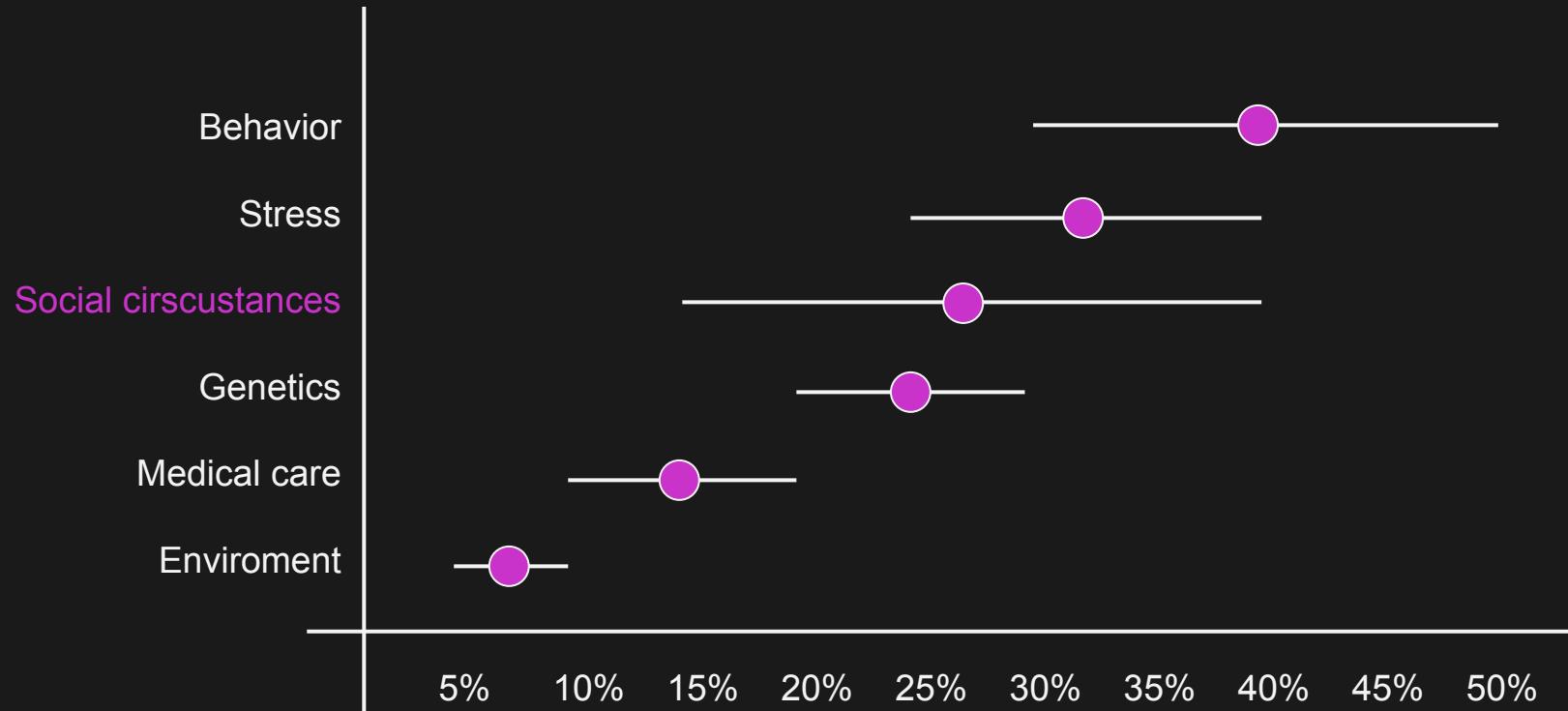
Rafael Nozal Cañadas

BiN-Tek Seminar / Sommarøy 2022





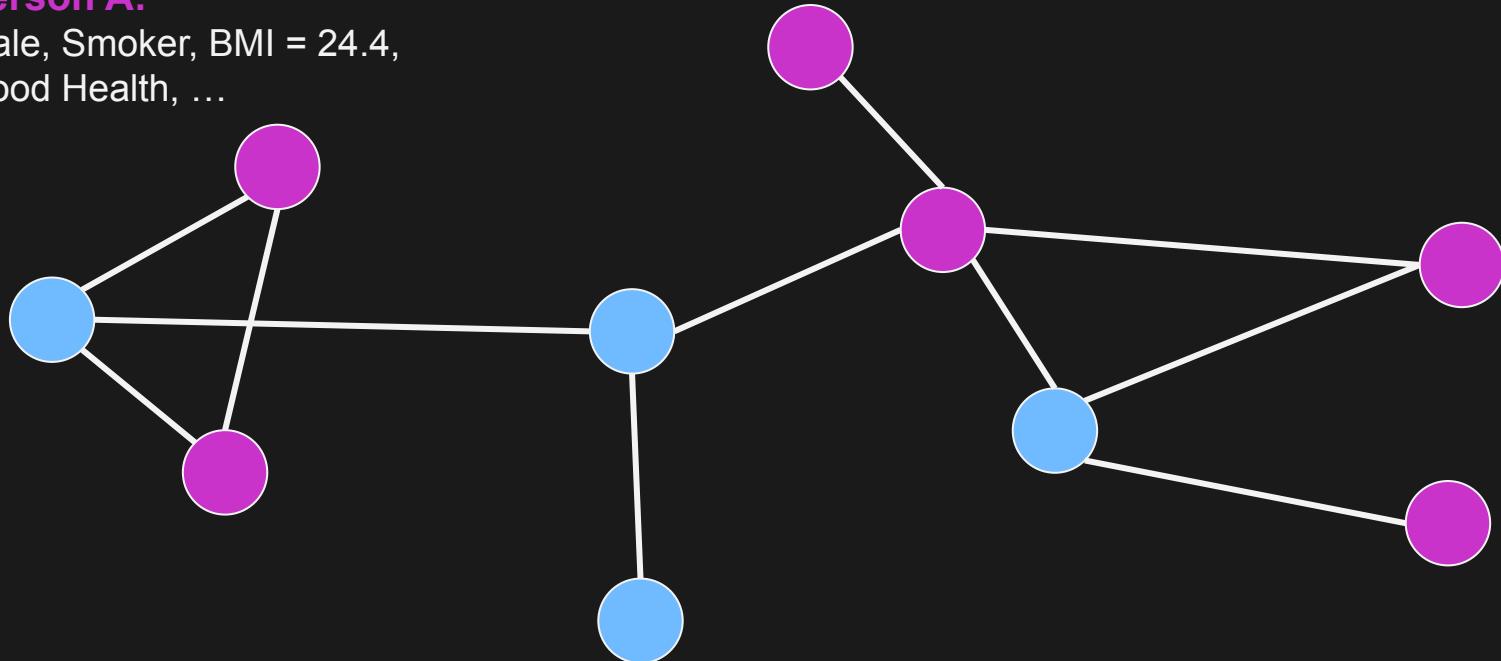
CHAPTER 0: Introduction



*"Health Policy Brief: The Relative Contribution of Multiple Determinants to Health Outcomes,"
Health Affairs, August 21, 2014*

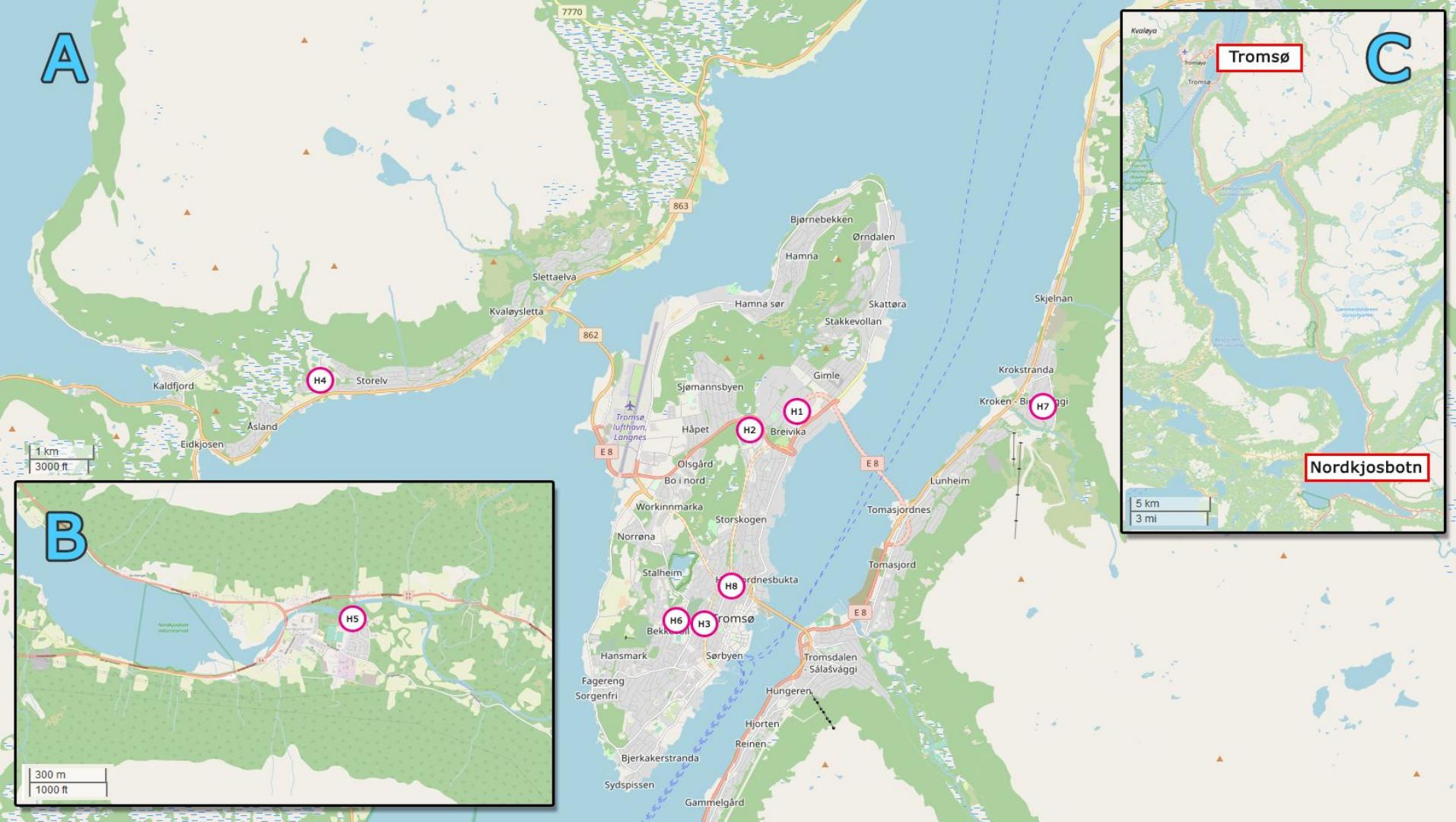
Person A:

Male, Smoker, BMI = 24.4,
Good Health, ...



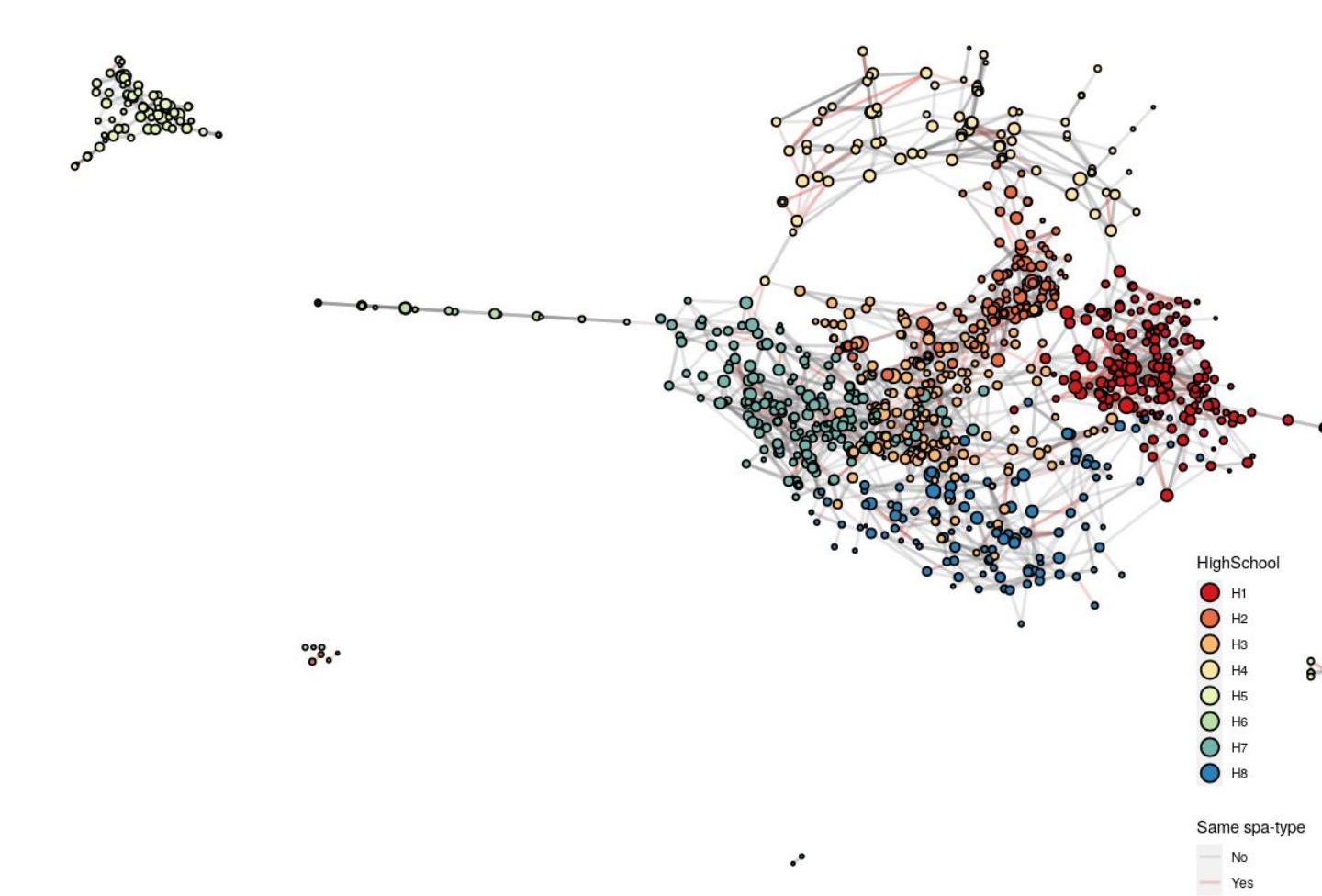
Person B:

Female, Non-Smoker, BMI = 22.1,
Good Health, ...



School network, highschools in nodes and same spa-type in edges.

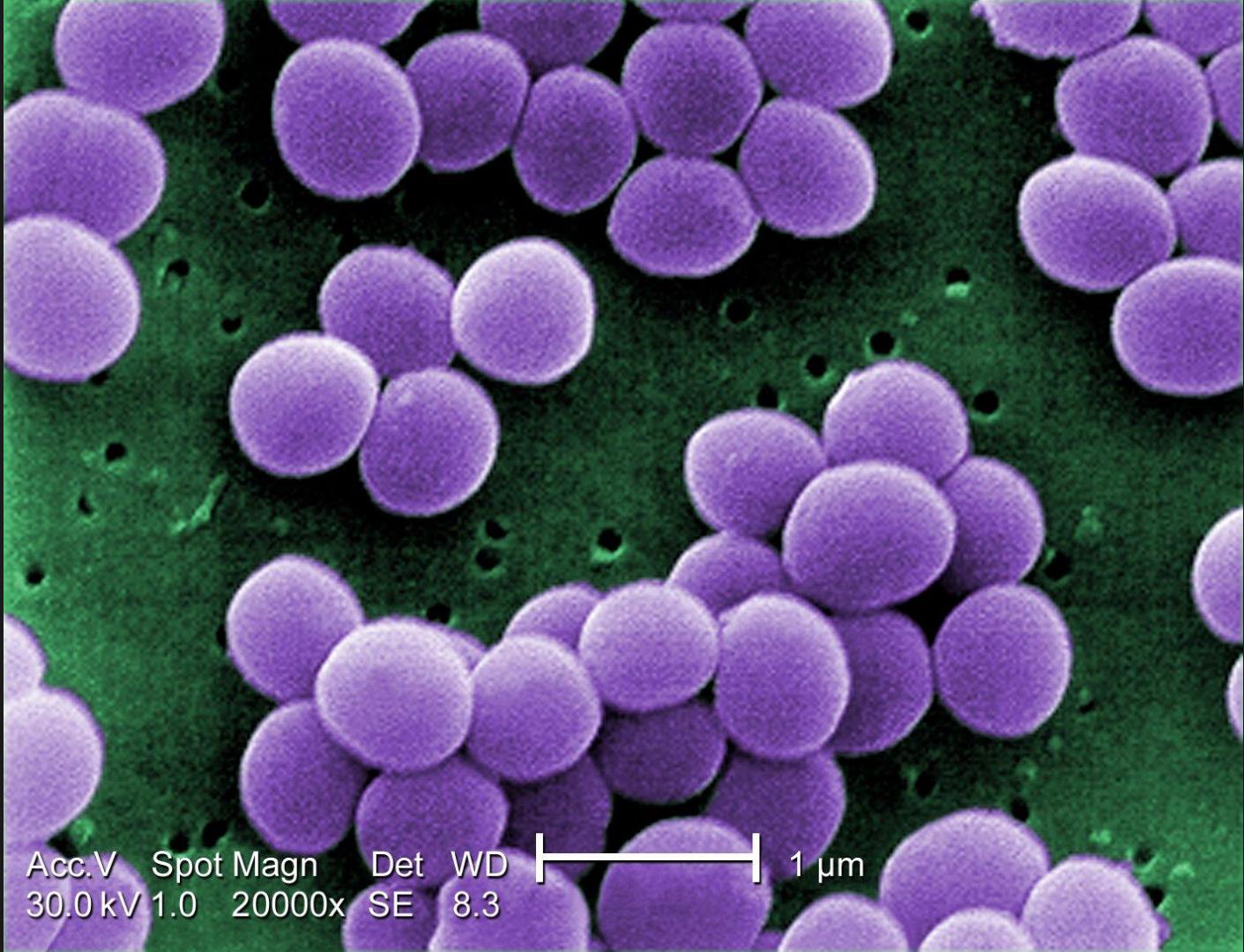
MDS layout, with isolated nodes hidden (n = 21)



Semi/Full autotomatic analysis:

- R is old arquitecture, make R more into C++
- Don't waste time micromanaging statistics
- Testing ontology annotation
- Clean and understand the data VERY well

CHAPTER 1: STAPH infections



Acc.V Spot Magn Det WD
30.0 kV 1.0 20000x SE 8.3

Pros

Harmless on the skin

Pros

Harmless on the skin

Cons

Pneumonia, Meningitis, Flesh eating disease!

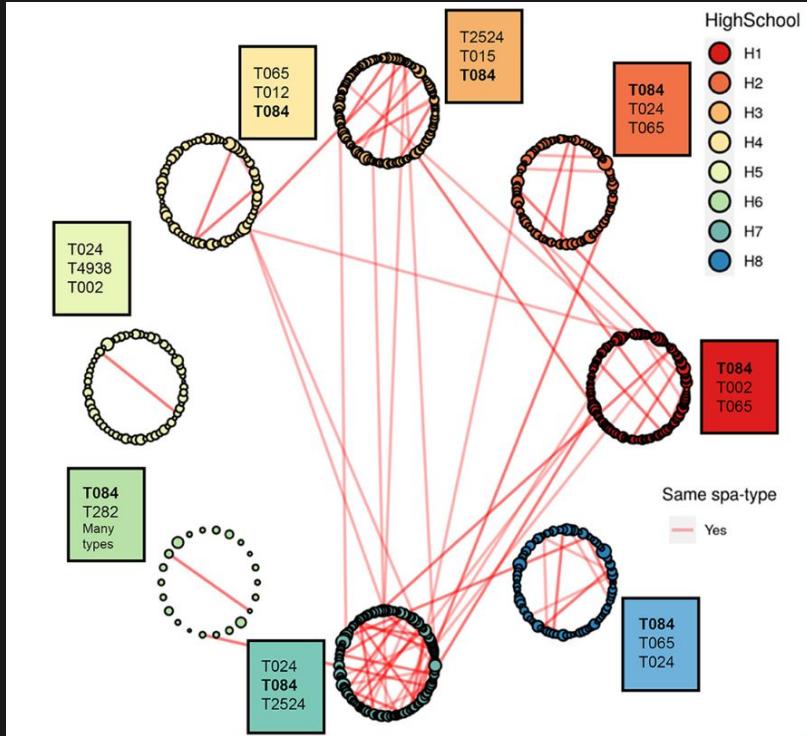
SAB deaths > AIDS, tuberculosis, and viral hepatitis combined.

20% hospital infections,

Immunoevasion, β -lactam antibiotic resistance. No vaccine.

...

Popular SPA-type profiles by Highschool



Adapted multivariable linear autocorrelation model.

	Estimate	SE	P-value
p	0.048	0.011	****
Sex	-0.048	0.028	****
Study program	0.043	0.022	*
BMI	0.107	0.018	****
Smoke	-0.012	0.027	0.65
Snuff	-0.001	0.020	0.96
Alcohol	0.038	0.021	0.07
Physical activity	0.033	0.012	***

CHAPTER 2:

Obesity and inflammation

The NEW ENGLAND JOURNAL of MEDICINE

SPECIAL ARTICLE

The Spread of Obesity in a Large Social Network over 32 Years

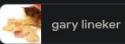
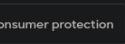
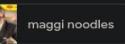
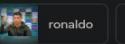
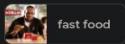
Nicholas A. Christakis, M.D., Ph.D., M.P.H., and James H. Fowler, Ph.D.

The NEW ENGLAND JOURNAL of MEDICINE

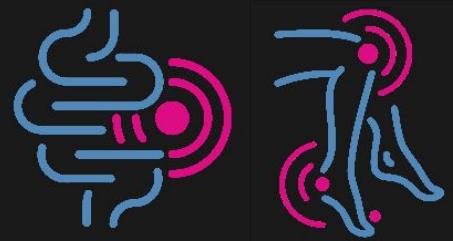
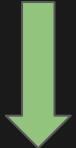
SPECIAL ARTICLE

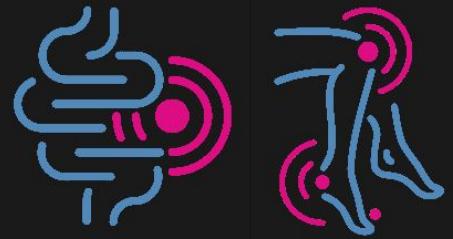
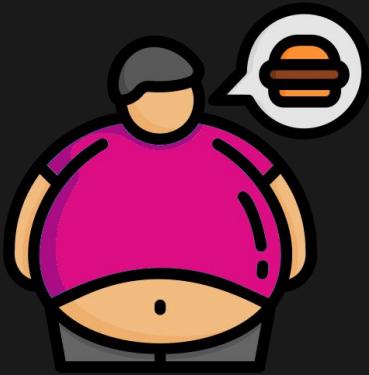
Your friends make you fat

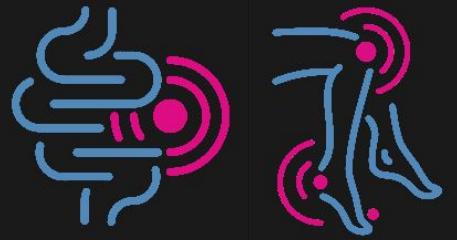
Nicholas A. Christakis, M.D., Ph.D., M.P.H., and James H. Fowler, Ph.D.

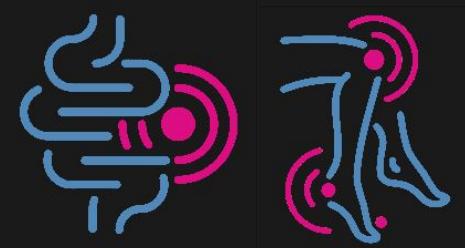
Celebrity Endorsements Sling Unhealthy ...
insteading.comAthletes in junk food ads | CBC.ca
cbc.caTeen fave musicians endorse mostly junk ...
cnn.comCelebrity Endorsements Sling Unhealthy ...
insteading.comThe Era of the Celebrity...
nytimes.comCelebrity endorsements 'effective' at ...
foodnavigator.com7 Unhealthy Brands Endorsed by Athletes
businessinsider.comTeen fave musicians endorse mostly junk ...
cnn.comPaid Millions to Endorse Junk Food ...
sciencealert.comcelebrities hawking fast food and soda ...
globalnews.caFast Food and Junk Fo...
marieclaire.comWeirdest Celeb Food and Drink Endorsemen...
spoonuniversity.comFood promoted by music stars usually ...
foodmag.com.auendorse junk food ...
indiantelevisions.comCelebrity Endorse...
insteading.comJunk Food Celebrities
thedoctortwillseeyounow.comFast Food and Junk Food Commercials ...
marieclaire.comIndian celebs have endorsed junk food ...
therprint.instop endorsing 'unhealthy' food ...
folomojo.com5 Random Celebrity Junk Foods - View ...
viewthevibe.comCelebs Who Endorsed Unhealthy Foods
therecipe.compromote junk foods and drinks ...
foodnavigator-asia.comUK ban on junk food ads on TV ...
theguardian.comCelebrities Endorse Mostly Unhealthy Foods
healthline.comTeens and parents urged to challenge ...
cbc.caAthletes Support Obesity - IMC Haw...
imcclass.comPepsi and Celebrity Endors...
abimbolatutu.wordpress.com



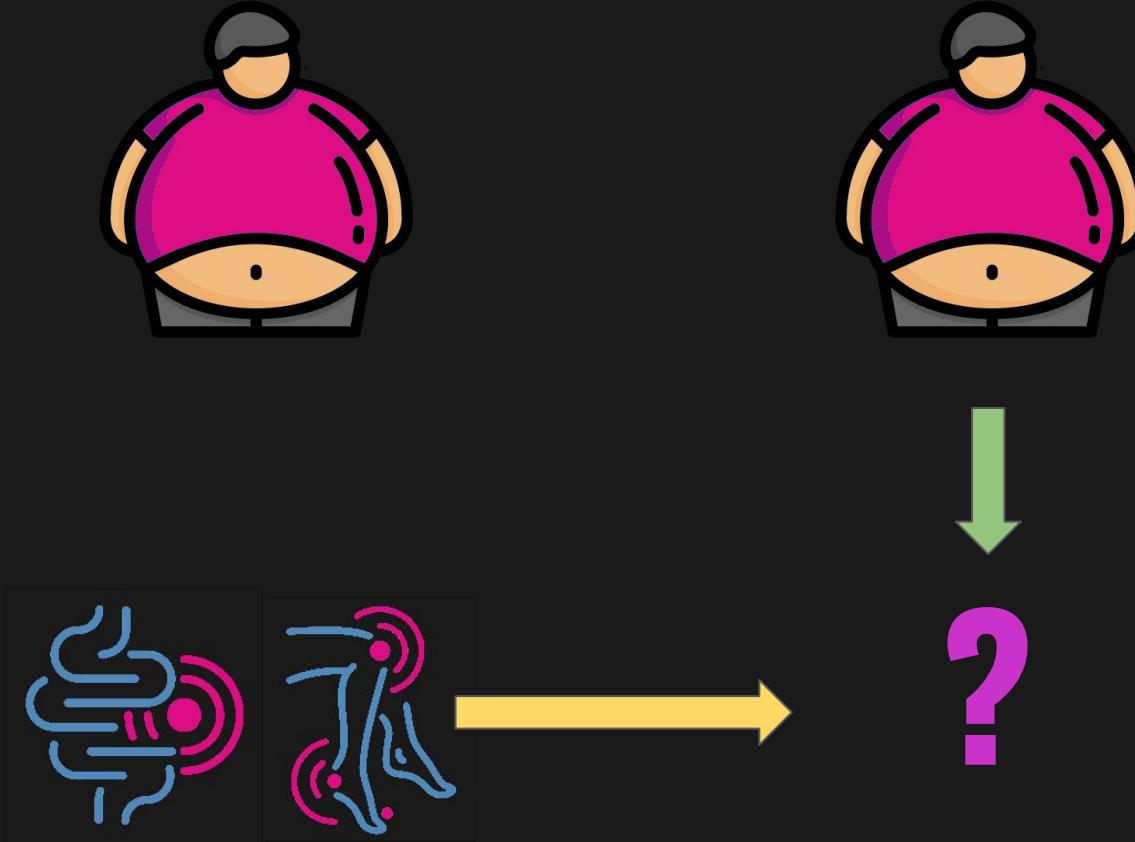








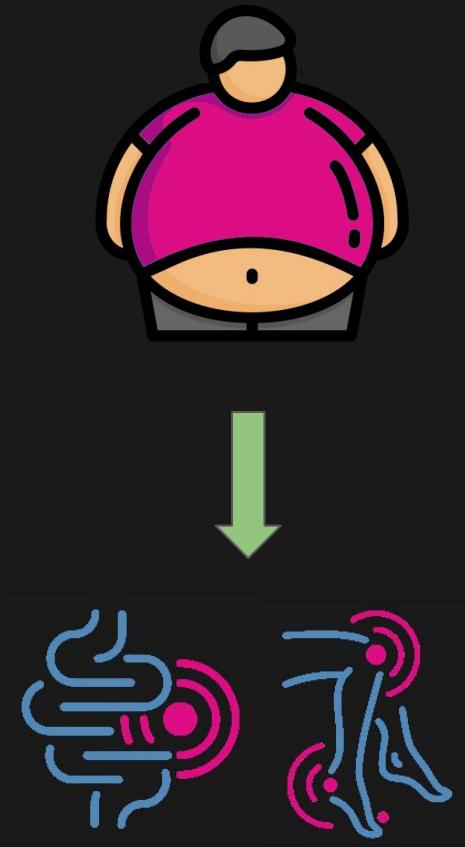
?





Network	Total Relationships	Equal Relationships	MIN	Q1	Median	Average	Q3	MAX	SD	p-value
Overall	3767	2043	1691	1846	1890.5	1892.661	1936	2140	69.7	0.016
Physical	2823	1584	1240	1379	1417	1417.441	1458	1590	57.13	0.002
School	2979	1590	1318	1459	1497	1497.958	1538	1666	56.23	0.051
Sports	598	415	233	288	301	301.938	316	371	20.78	0
Home	1247	722	530	605	627	626.908	648	738	31	0.001
Other	1095	612	450	531	551	550.655	570	648	28.63	0.016

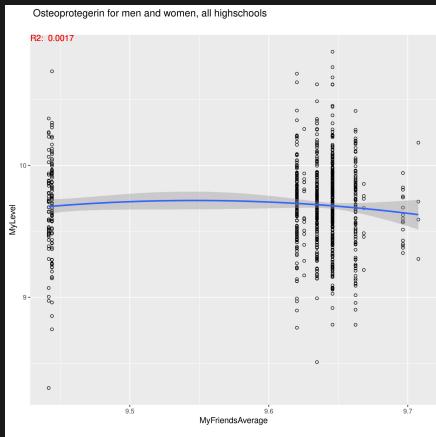
Table 5: Simulated networks ($n=1000$) same-to-same relationships against the real network same-to-same relationships. All simulated network shows bias of BMI spread in the real network.



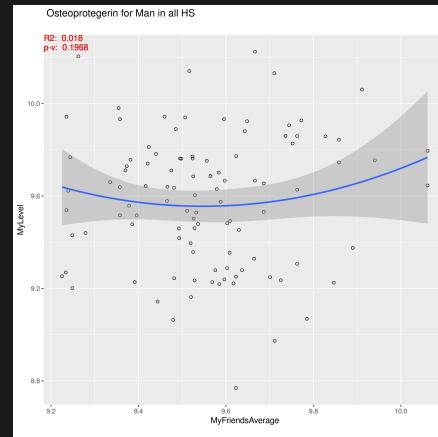
Protein	Waist	Hip	Height	Weight	BMI	HR	SYSBP	DIABP
Caspase-8	*	***	ns	***	***	ns	ns	ns
C-C motif chemokine 3	*	ns	ns	*	ns	ns	ns	ns
CUB domain-containing protein 1	****	****	ns	****	****	ns	ns	ns
Macrophage colony-stimulating factor 1	****	***	ns	**	**	ns	ns	ns
Delta and Notch-like epidermal growth factor-related receptor	ns	ns	ns	*	*	ns	ns	ns
Fibroblast growth factor 21	*	*	ns	ns	*	ns	ns	ns
Hepatocyte growth factor	****	***	ns	**	****	ns	ns	ns
Interleukin-10 receptor subunit beta	****	*	ns	**	**	ns	ns	ns
Interleukin-18	**	*	ns	*	**	ns	ns	ns
Interleukin-18 receptor 1	****	***	ns	***	****	ns	ns	ns
Interleukin-2	*	ns	ns	ns	ns	ns	ns	ns
Interleukin-6	****	****	ns	****	****	ns	ns	ns
Interleukin-7	**	**	ns	**	*	ns	ns	ns
Monocyte chemoattractant protein 3	****	****	ns	****	****	ns	ns	ns
Monocyte chemoattractant protein 4	*	ns	ns	ns	*	ns	ns	ns
Latency-associated peptide transforming growth factor beta-1	*	*	ns	ns	ns	ns	ns	ns
TNF-related apoptosis-inducing ligand	**	*	ns	ns	*	ns	ns	ns
TNF-related activation-induced cytokine	*	**	ns	*	ns	ns	ns	ns
Vascular endothelial growth factor A	**	*	ns	*	***	ns	ns	ns

Table 6: Biomarkers that are statistically significant with respect the antropometry variables in women, after applying Bonferroni correction

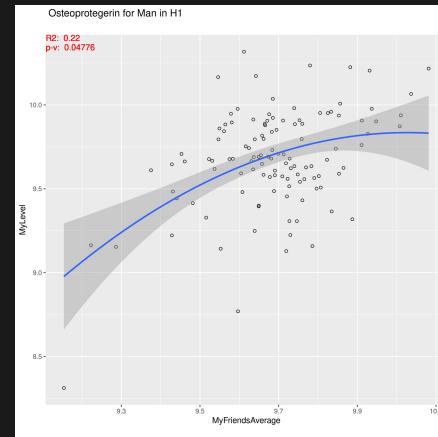
No stratification



Stratified for sex only



Stratified sex and highschool

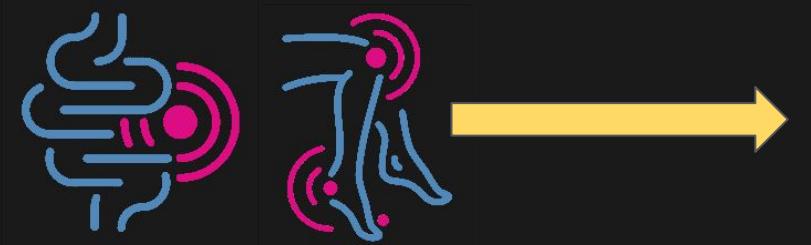


Protein	Men	Women
Adenosine Deaminase	1.12	1.03
Artemin	0.98	1.04
Axin-1	1.06	1.02
Brain-derived neurotrophic factor	1.06	1.14
Beta-nerve growth factor	1.29	0.91
Caspase-8	1.01	1.04
Eotaxin	0.94	1.04
C-C motif chemokine 19	1	0.96
C-C motif chemokine 20	1.09	0.97
C-C motif chemokine 23	1.06	1.1
C-C motif chemokine 25	1.04	1.02
C-C motif chemokine 28	0.76	0.84
C-C motif chemokine 3	1	0.96
C-C motif chemokine 4	0.98	1.02
Natural killer cell receptor 2B4	1.02	1
CD40L receptor	1.03	1.04
Tcell surface glycoprotein CD5	1.09	1.07
T cell surface glycoprotein CD6 isoform	1.04	0.97
CUB domain-containing protein 1	1.04	1.14
Macrophage colony-stimulating factor 1	1.01	1.13
Cystatin D	1.04	0.97
Fractalkine	1.12	1.03

Enemies Mean Squared NPX Distance

Friends Mean Squared NPX Distance

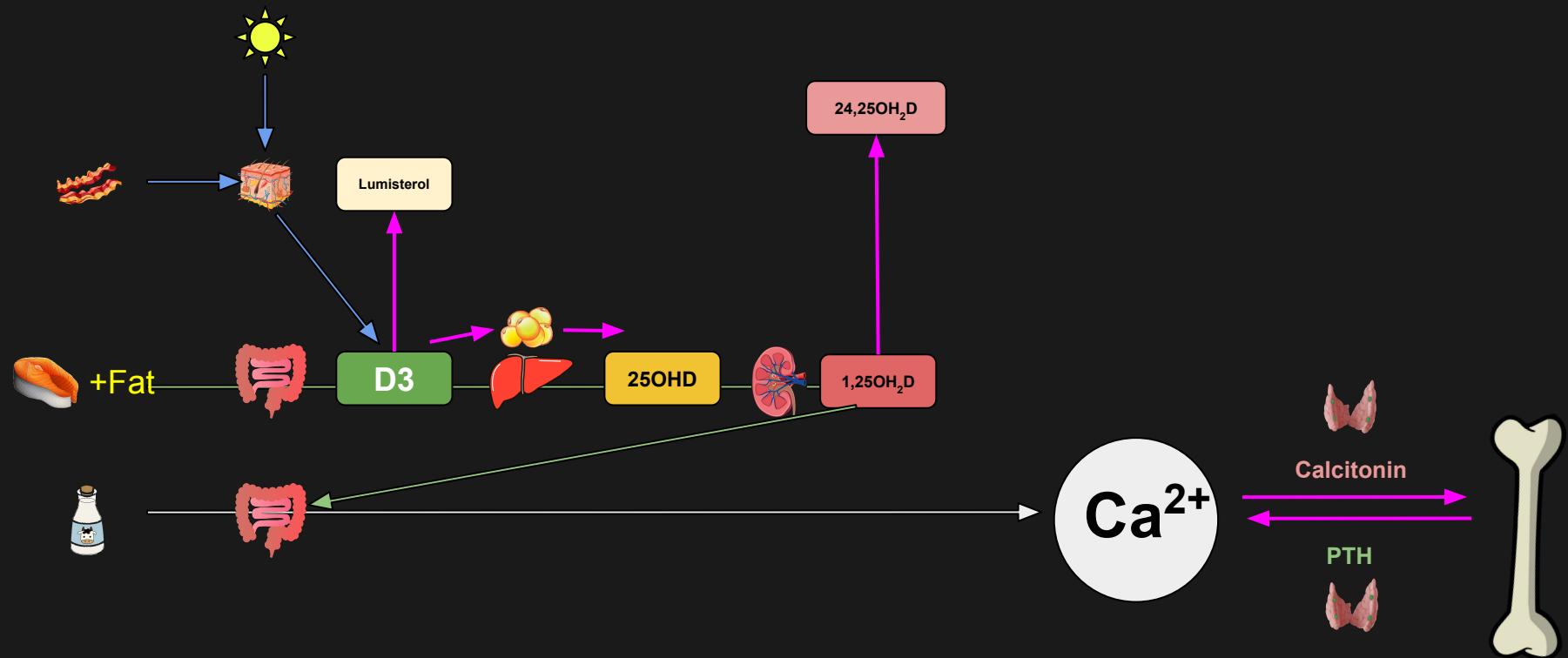
(NPX = Normalized Protein eXpression)

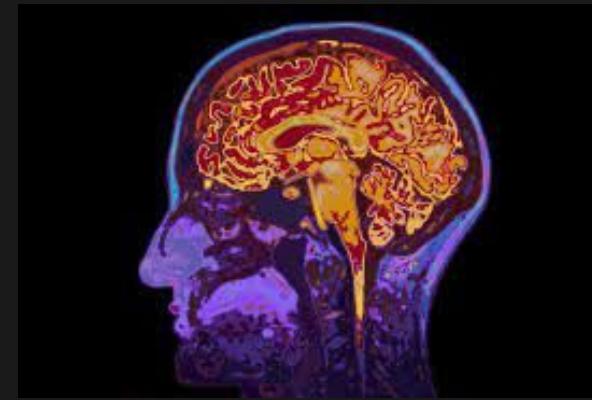


- Too much data to explore
- No medics to look at it :(

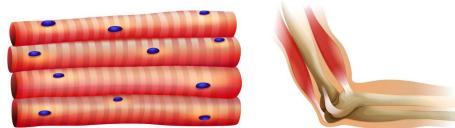
CHAPTER 3:

Vitamin D

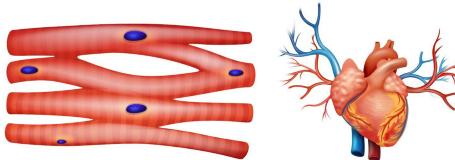




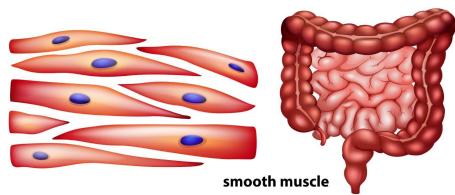
Types of Muscle Cells



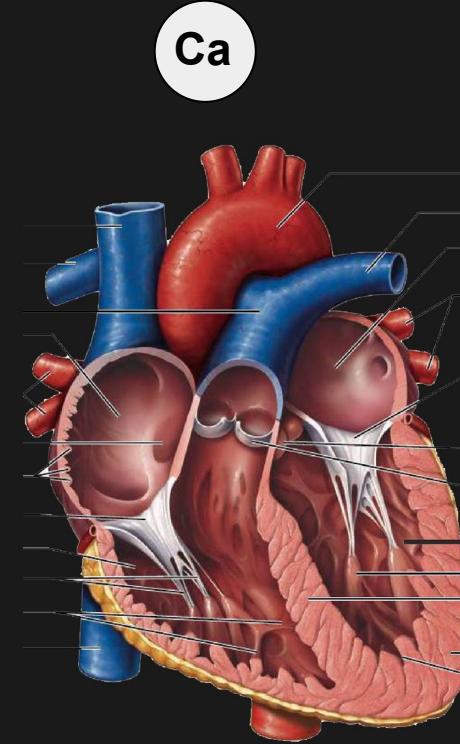
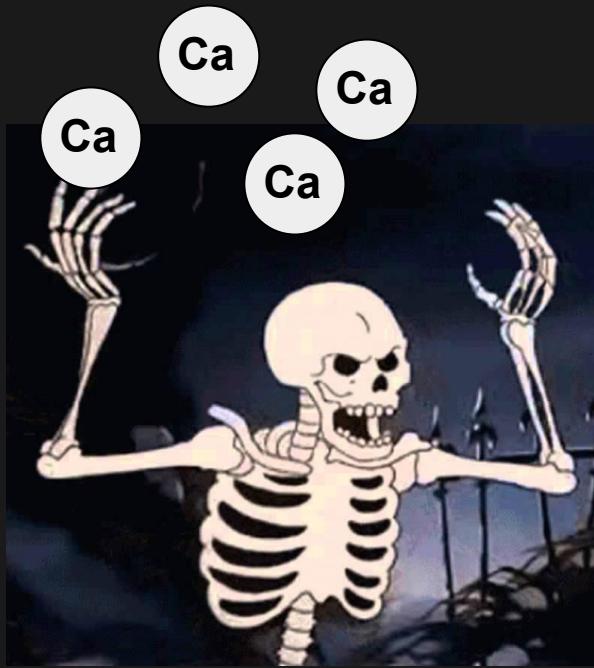
skeletal muscle

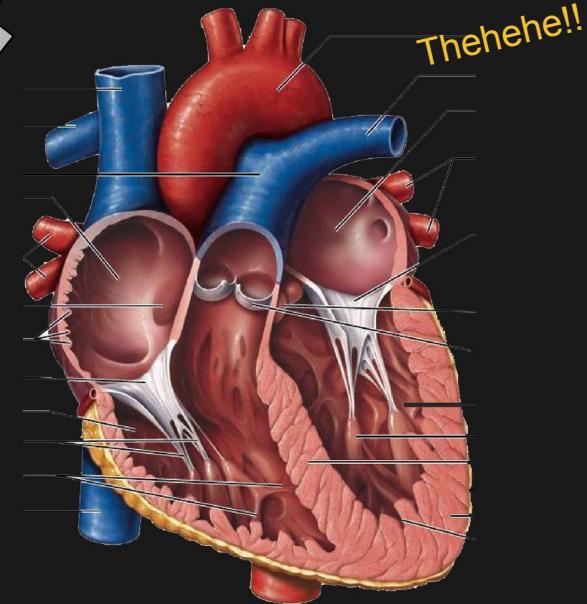
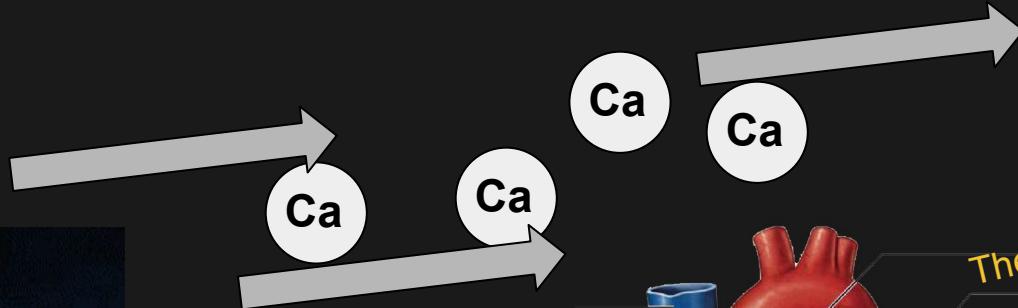
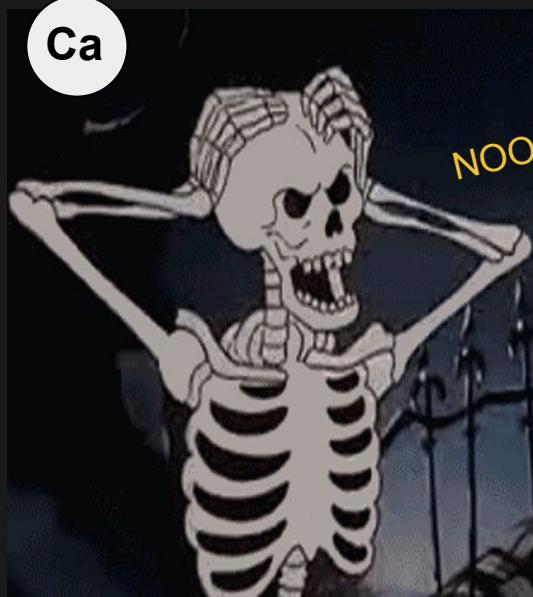


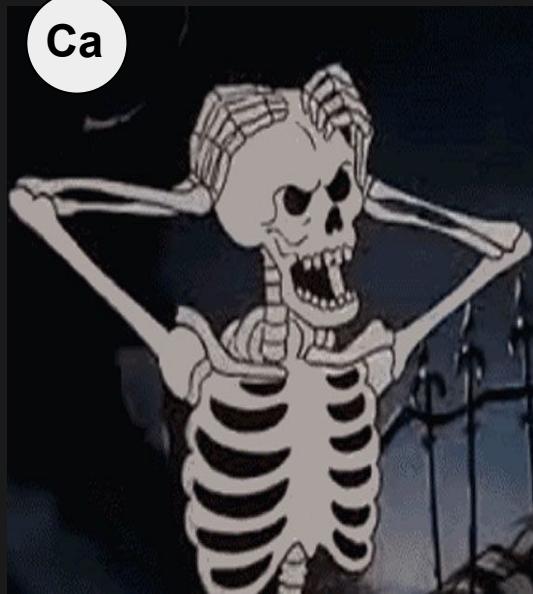
cardiac muscle



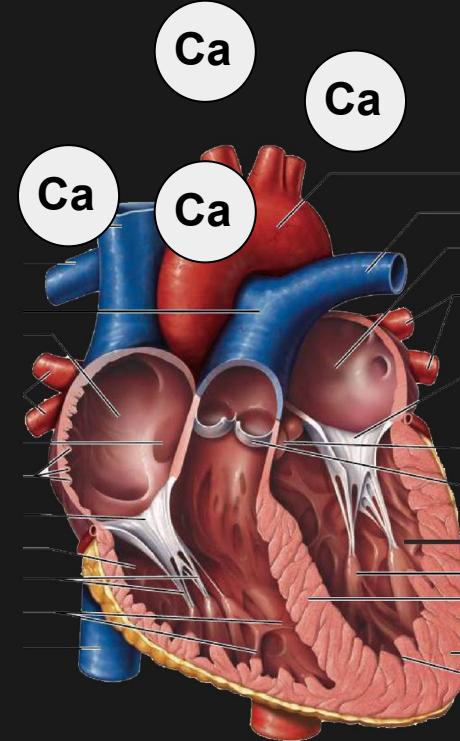
smooth muscle





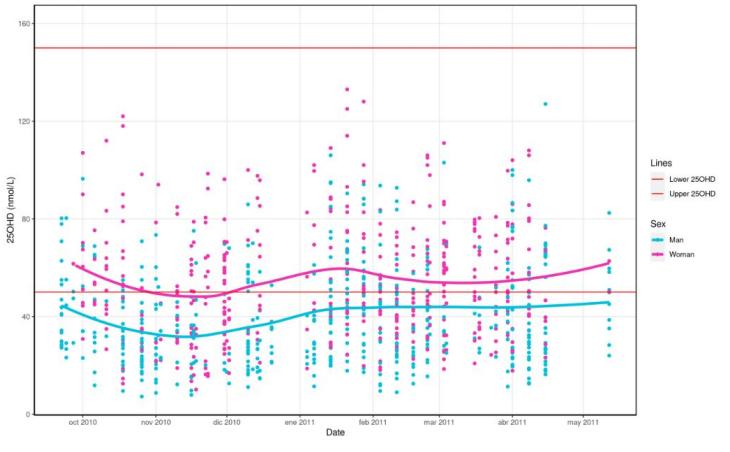


Cons: Rickets
Osteoporosis
Artrosis

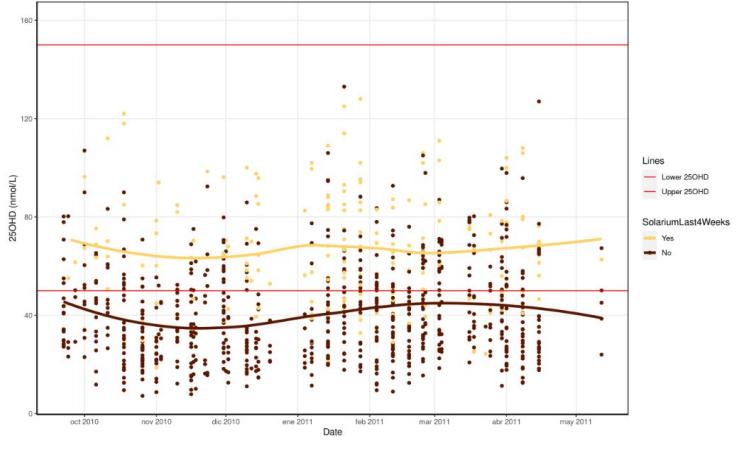


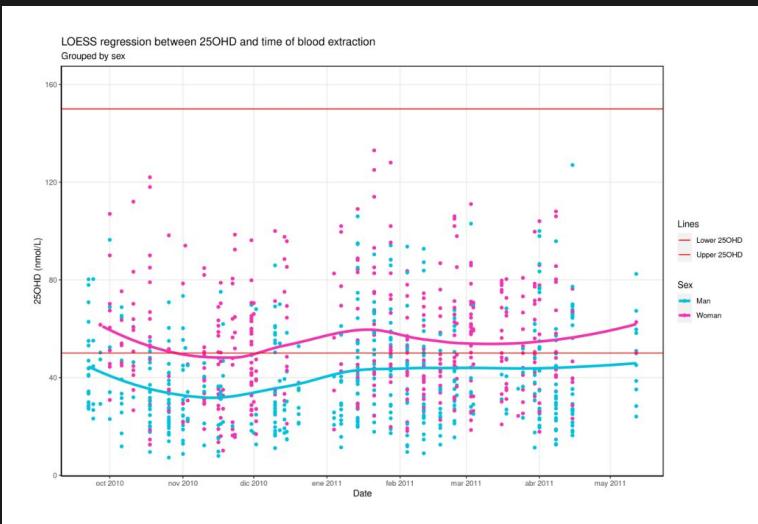
Pros: Working heart

LOESS regression between 25OHD and time of blood extraction
Grouped by sex

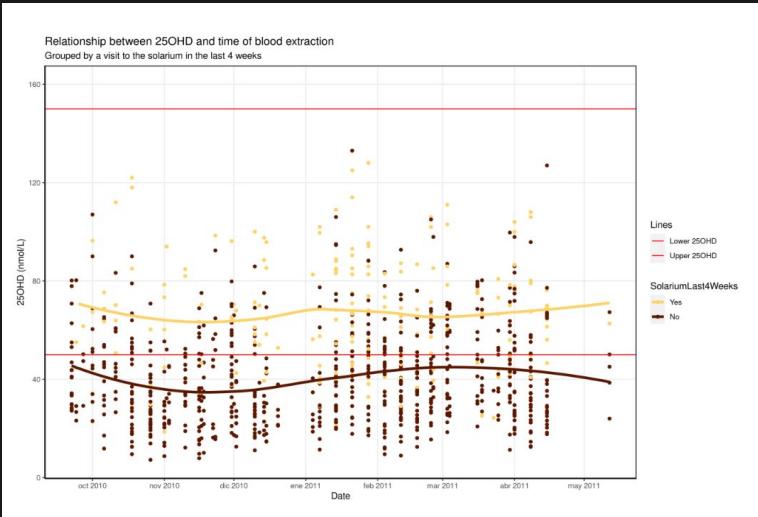


Relationship between 25OHD and time of blood extraction
Grouped by a visit to the solarium in the last 4 weeks





People are friends with people with same solarium habits. ($p_v = 0.008$)



Women do influence other women into going to the solarium. ($p_v = 0.0002$)

Very few friendships between men and women to tell cross influence.

DON'T GO TO THE SOLARIUM

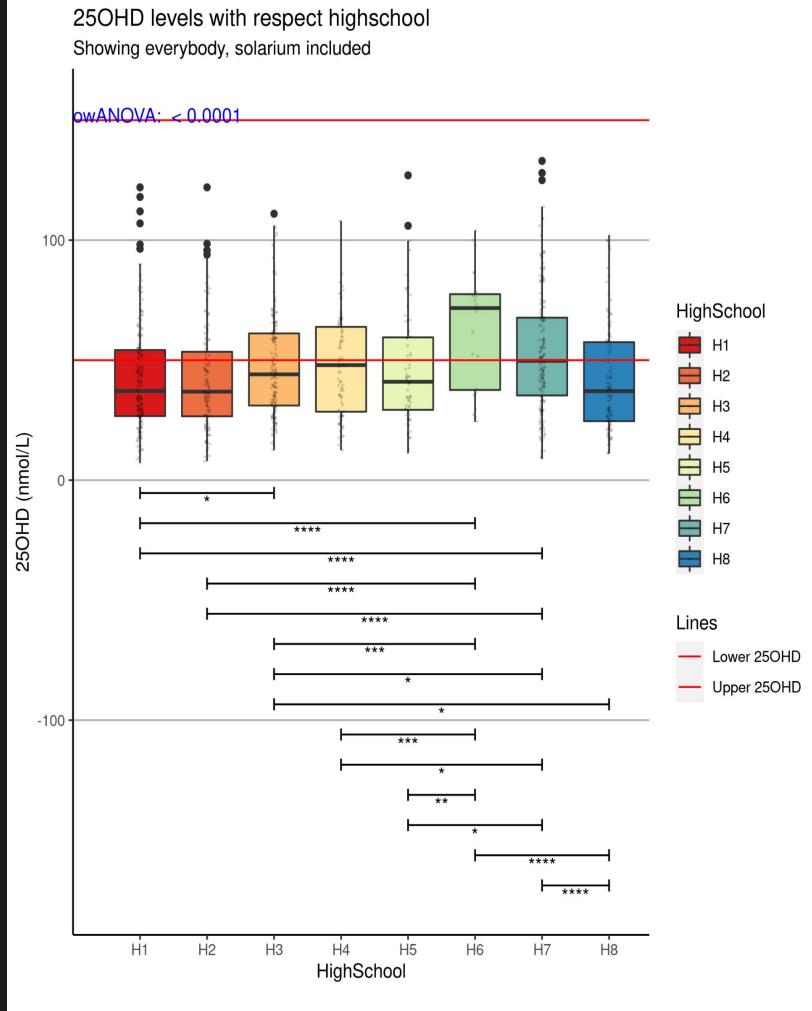
*Prevalence of Indoor Tanning Among Teenagers in Norway Before and After Enforcement of Ban for Ages Under 18 Years
Short communication, Trude E. ROBSAHM, Jo S. STENEHJEM, Leon A.M. BERGE1 and Marit B. VEIERØD*

*(NOWAC) Association of Lifetime Indoor Tanning and Subsequent Risk of Cutaneous Squamous Cell Carcinoma
JAMA Dermatology*

25OHD levels with respect highschool

Showing everybody, solarium included

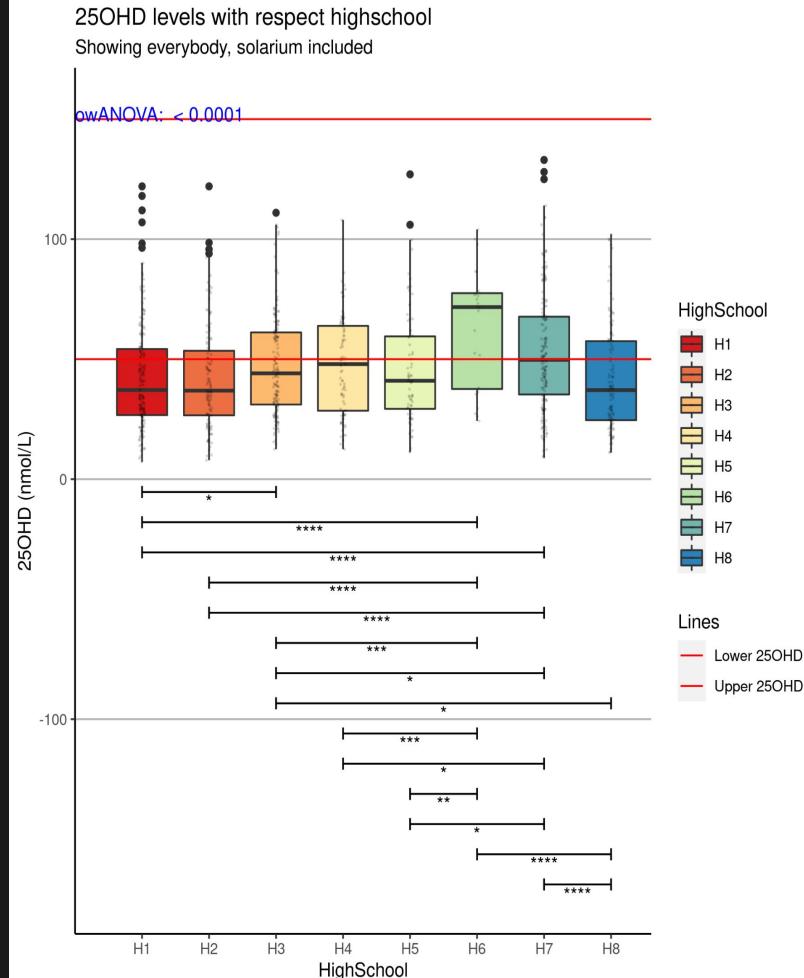
swANOVA: < 0.0001



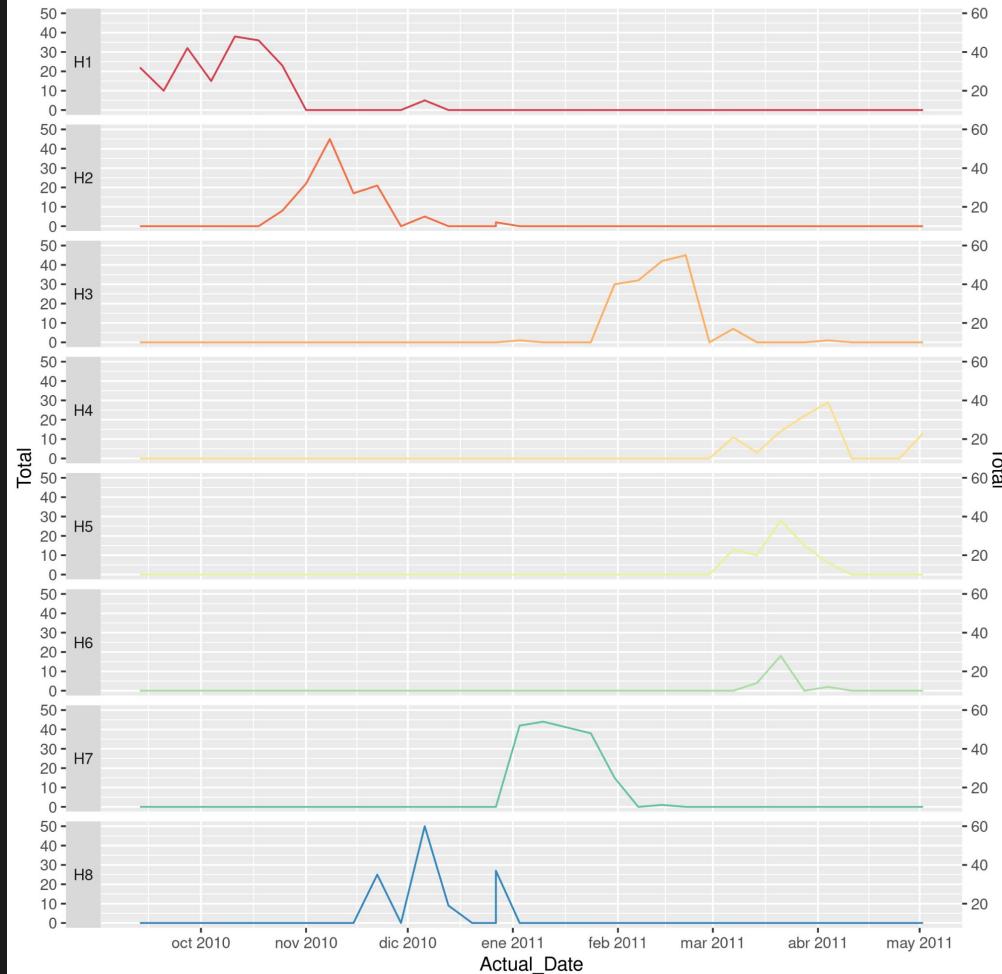
25OHD levels with respect highschool

Showing everybody, solarium included

ANOVA: < 0.0001

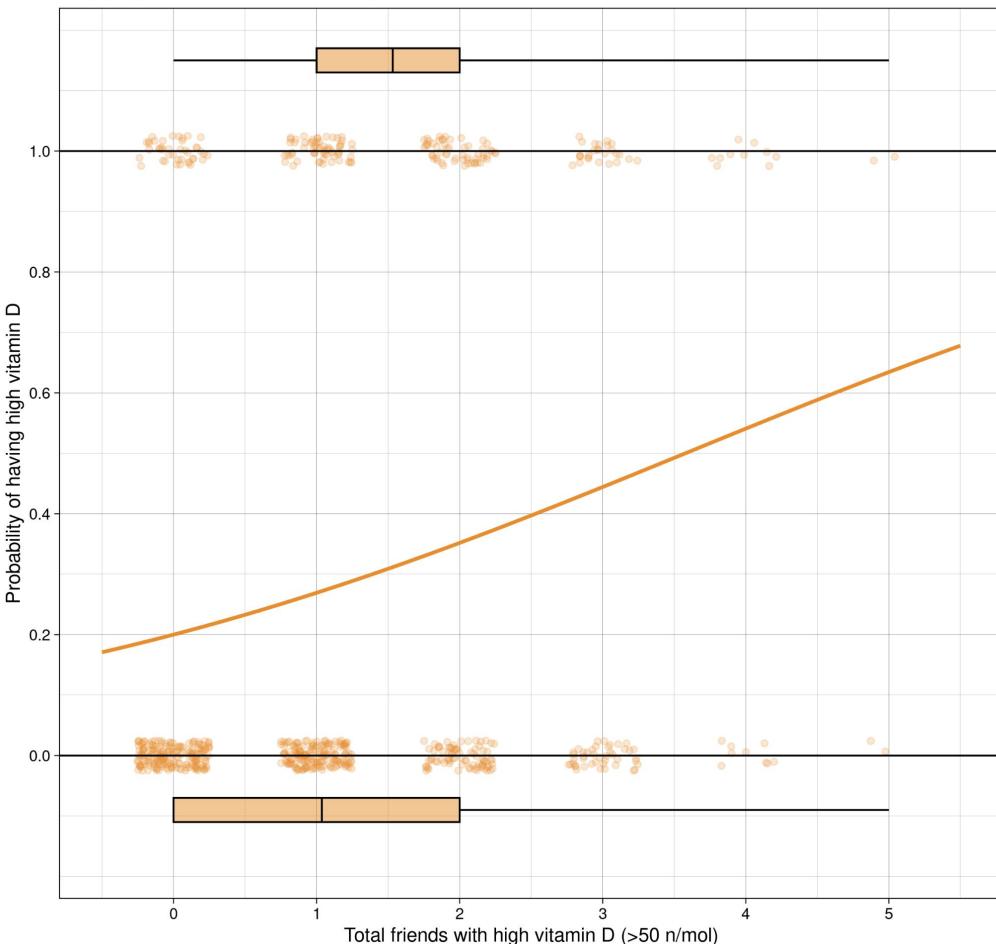


Time of the year for blood extraction for each high-school



Logit Regression between total high vitamin D friends and probability of high vitamin D status

Only people with friends are shown, p-value < 0.0001

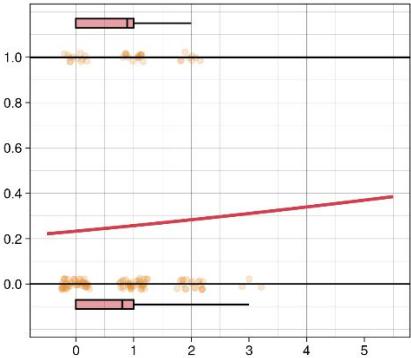


Logistic regression shows high vitamin D correlation levels between friends, meaning if you have friends with high (or low) vitamin D, you are likely to share their levels.

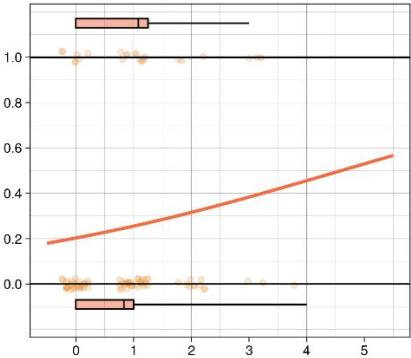
This is true even after correcting for people who don't go to the solarium.

χ^2 analysis also show vitamin D levels correlate (pv: 0.003) with high-schools independently of time of the year during blood extraction.

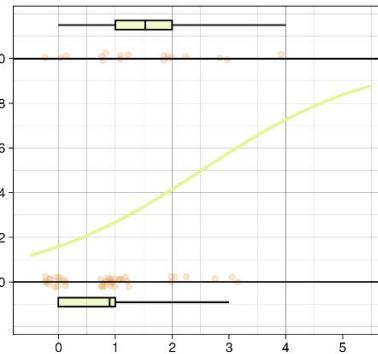
H1 and logit for vitamin D status
p-value = 0.5797



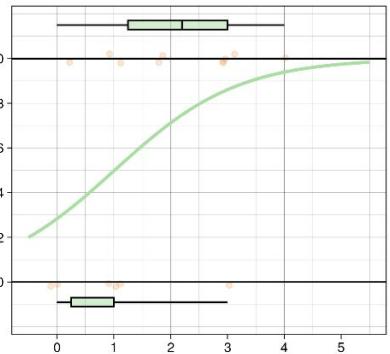
H2 and logit for vitamin D status
p-value = 0.2395



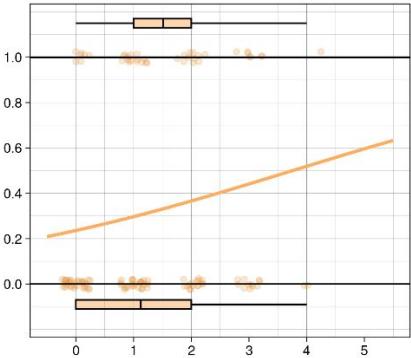
H5 and logit for vitamin D status
p-value = 0.025



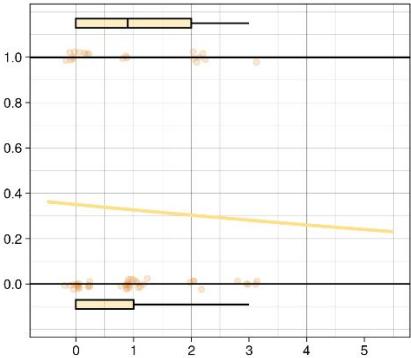
H6 and logit for vitamin D status
p-value = 0.0697



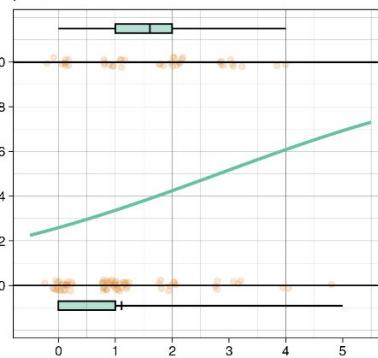
H3 and logit for vitamin D status
p-value = 0.0651



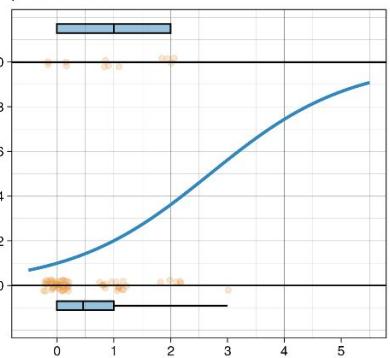
H4 and logit for vitamin D status
p-value = 0.7124



H7 and logit for vitamin D status
p-value = 0.0282

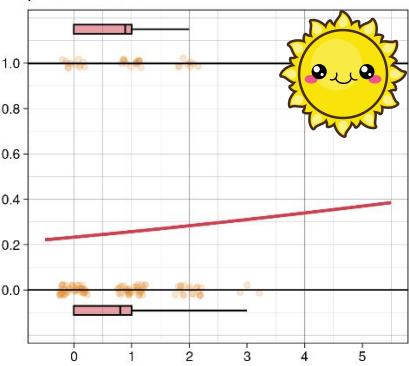


H8 and logit for vitamin D status
p-value = 0.024



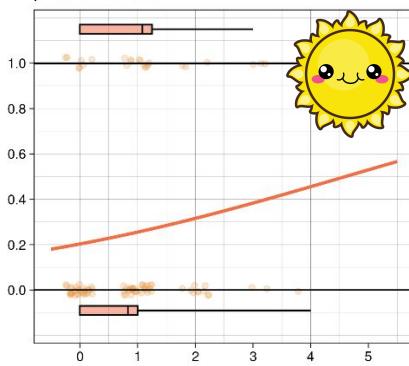
H1 and logit for vitamin D status

p-value = 0.5797



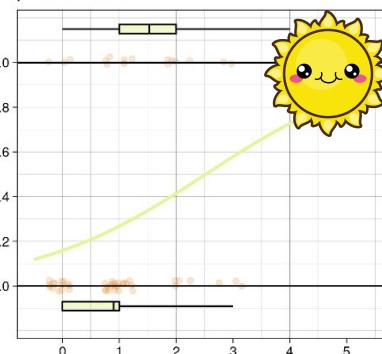
H2 and logit for vitamin D status

p-value = 0.2395



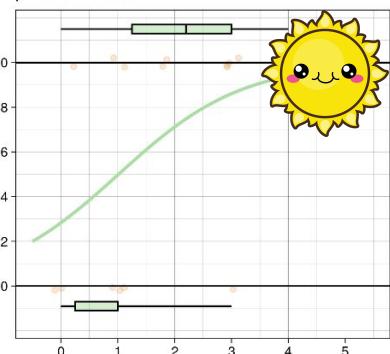
H5 and logit for vitamin D status

p-value = 0.025



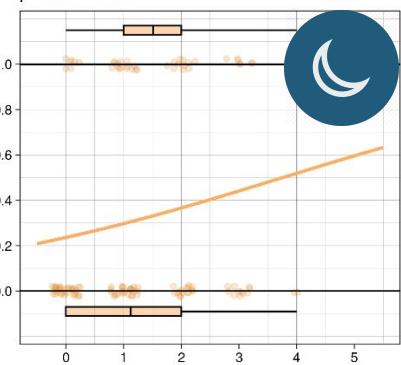
H6 and logit for vitamin D status

p-value = 0.0697



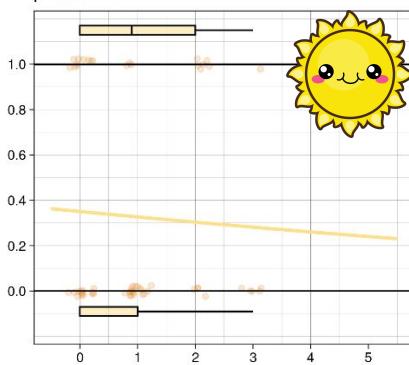
H3 and logit for vitamin D status

p-value = 0.0651



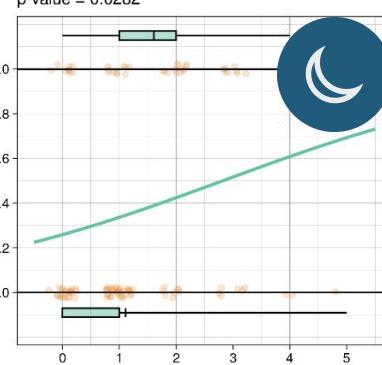
H4 and logit for vitamin D status

p-value = 0.7124



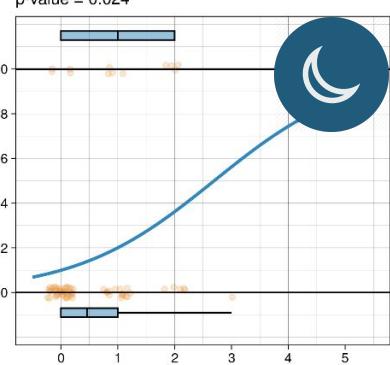
H7 and logit for vitamin D status

p-value = 0.0282

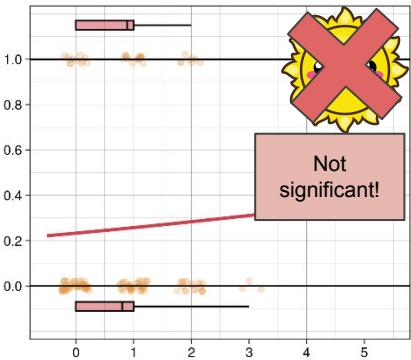


H8 and logit for vitamin D status

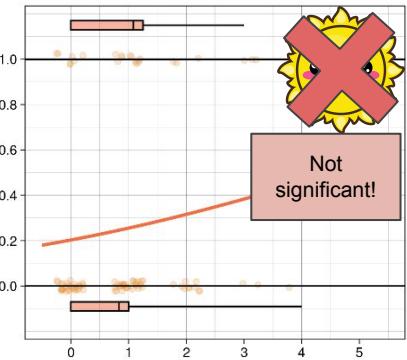
p-value = 0.024



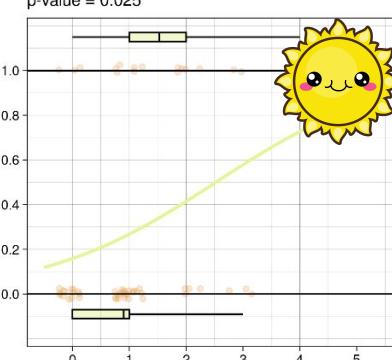
H1 and logit for vitamin D status
p-value = 0.5797



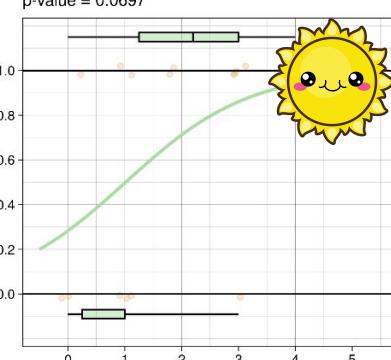
H2 and logit for vitamin D status
p-value = 0.2395



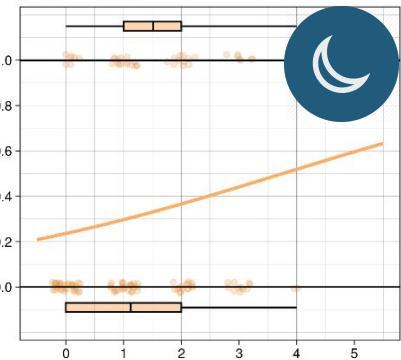
H5 and logit for vitamin D status
p-value = 0.025



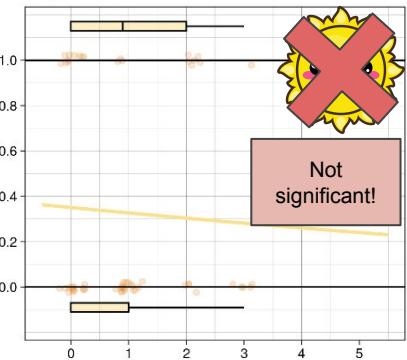
H6 and logit for vitamin D status
p-value = 0.0697



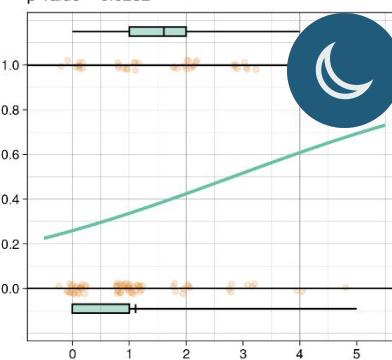
H3 and logit for vitamin D status
p-value = 0.0651



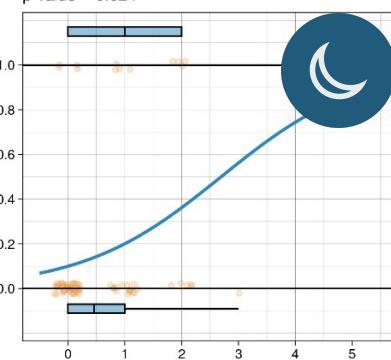
H4 and logit for vitamin D status
p-value = 0.7124



H7 and logit for vitamin D status
p-value = 0.0282



H8 and logit for vitamin D status
p-value = 0.024



CHAPTER 4: Extras

Title	Journal	Published	DOI
An introduction to network analysis for studies of medication use	[Q1] Research in Social and Administrative Pharmacy	2021 December	doi.org/10.1016/j.sapharm.2021.06.021
TTT-SiZer: A graphic tool for aging trends recognition	[Q1] Reliability Engineering & System Safety	2020 October	10.1016/j.ress.2020.107010
Statistical supervised learning with engineering data: a case study of low frequency noise measured on semiconductor devices	[Q1] The International Journal of Advanced Manufacturing Technology	2022	10.1007/s00170-022-08949-z
Unsupervised and supervised learning for the reliability analysis of complex systems	rejected once, oh no! :(2022!	
Automated coding of Danish causes of death 1861-1911. How far can we get with string similarity?	Abstract only: Digital Humanities in the Nordic and Baltic countries (DHNB)	2023?	

CHAPTER 5: Improving the data

- a) Data access
- b) Data cleaning
- c) Questionary design

- a) Data access
- b) Data cleaning
- c) Questionary design

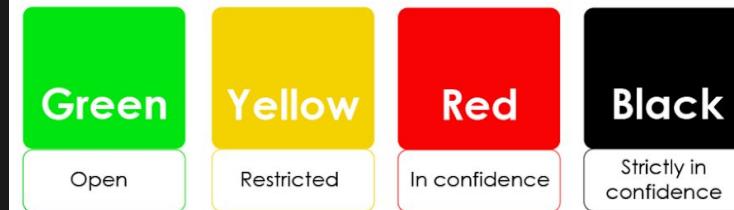
- a) Data access
- b) Data cleaning
- c) Questionary design

How to classify data and information

Norwegian

Table of Contents

- On ownership – how to decide which class to use
- Open or freely available (Green)
- Restricted (Yellow)
- In confidence (Red):
- Strictly in confidence (Black):



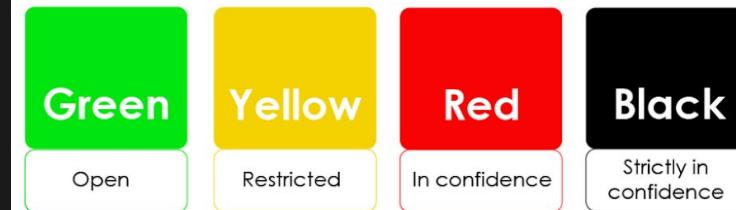
- a) Data access
- b) Data cleaning
- c) Questionary design

How to classify data and information

Norwegian

Table of Contents

- On ownership – how to decide which class to use
- Open or freely available (Green)
- Restricted (Yellow)
- In confidence (Red):
 - Strictly in confidence (Black):



Started June 2019

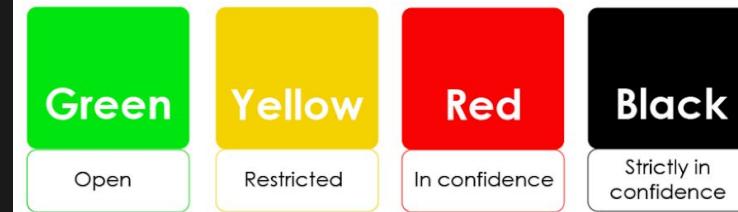
- a) Data access
- b) Data cleaning
- c) Questionary design

How to classify data and information

Norwegian

Table of Contents

- On ownership – how to decide which class to use
- Open or freely available (Green)
- Restricted (Yellow)
- In confidence (Red):
- Strictly in confidence (Black):

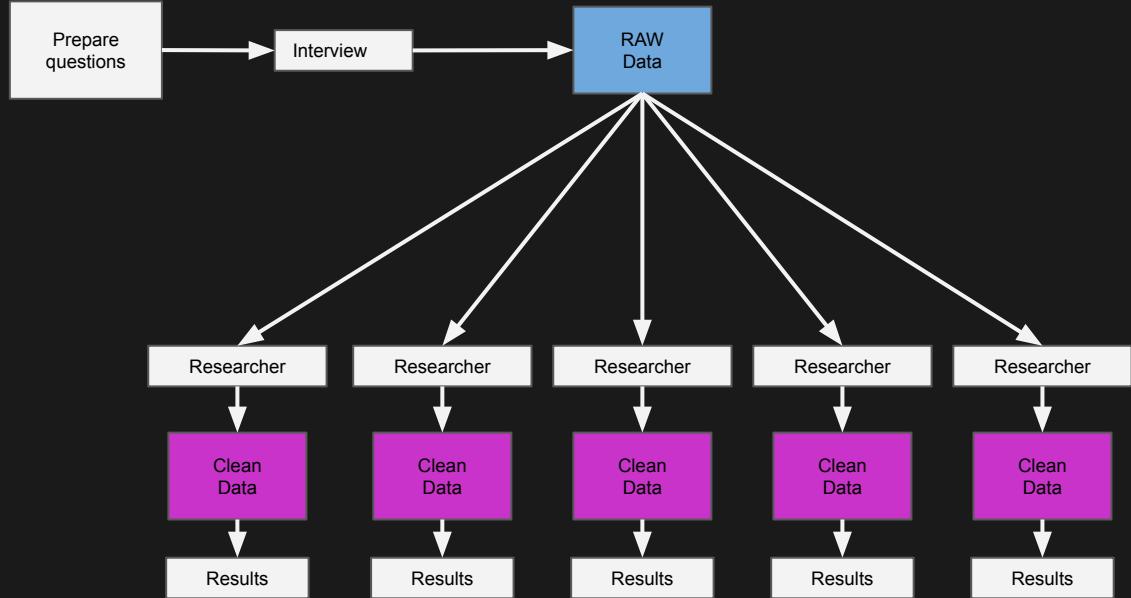


Started June 2019

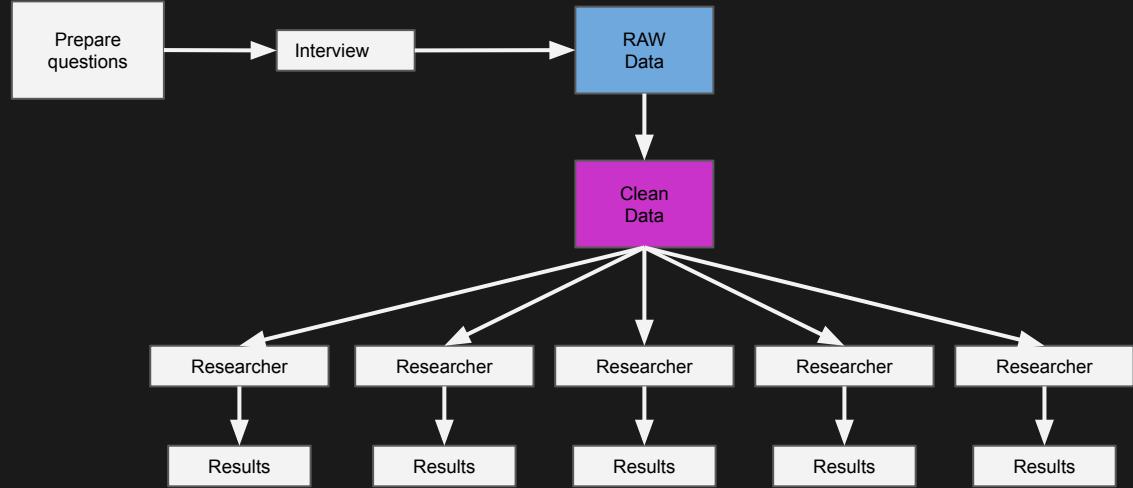
Network data on February 2020

First paper complete data on August 2021

- a) Data access
- b) Data cleaning
- c) Questionary design



- a) Data access
- b) Data cleaning
- c) Questionary design



- a) Data access
- b) Data cleaning
- c) Questionary design

“Other” diseases are not registered with proper ICD10

Questionary	Medical	ICD10
Skin diseases		
pustolosid palmoplantaris ev dermatitt. Beinhinnebetennelse i begge legger i tre år	Atopic dermatitis, unspecified	L20.9
pustolosid palmoplantaris ev dermatitt. Beinhinnebetennelse i begge legger i tre år	Osteomyelitis	M66.9
akne	Acne	L70
Eivblest	Urticaria, unspecified	L50.9
Kuldeallergi. Får utslett. Også når vått og fuktig .	Urticaria due to cold and heat	L50.2
vittigo, hudsykdom	Vitiligo	L80
Atopisk eksem	Atopic dermatitis, unspecified	L20.9
hudsykdom	Disorder of the skin and subcutaneous tissue, unspecified	L98.9
irritasjon, klee på øyne, diagnosert av lege	Pruritus, unspecified	L29.9
Digestive tract		
Morbus kronus	Crohns disease	K50.90
morbus crohn	Crohns disease	K50.90
Mb. Chron. Husker ikke når han fikk diagnosen. Var baby.	Crohns disease	K50.90
Magesår 2 ganger. Siste påvist i forrige uke. Starter beh.	Gastric ulcer	K25
Født med esofagusstresi (?). Ble behandlet i H.heim med blokking for 5 år siden."	Aesthesia of esophagus without fistula	Q39.0
Glutenintoleranse. Mangler enzymer. Oppdaget ved 15 års alderen	Celiac disease	K90.0
Kronisk pankreatitt	Other chronic pancreatitis	K86.1
Kidneys and genitals		
Kronisk urinveisinfeksjon.	Urinary tract infection, site not specified	N39.0
Dårlig funksjon av lukkemuskel ned til magen	Other disorders of urinary system	N39
PCOS	Polyzystic ovarian syndrome	E28.2
Kronisk smertesyndrom, type 2. (nyreproblematikk)	Chronic kidney disease, stage 2 (mild)	N18.2
General infectious diseases		
kysesyke (påvist i sept 2010, syk i mai	Infectious mononucleosis, unspecified without complication	B27.90
Autoimmune diseases		
Idiopatisk trombocytopenisk purpura	Immune thrombocytopenic purpura	D69.3
Lupus	Lupus erythematosus	L93
Autoimmun hepatitis	Autoimmune hepatitis	K75.4
Others		
tinnitus	Tinnitus	H93.1
ettere hørselhemmet	Unspecified sensorineural hearing loss	H80.5
(name redacted) har et syndrom exfraglisysyndrom (usikker hvordan det skrives) ¹	Fragile X chromosome	Q99.2
har i ett år hatt en tumor i høyre lår som skal opereres	Neoplasm of unspecified behavior of bone, soft tissue, and skin	D49.2
Hårravfall (husker ikke diagnose)	Nonscarring hair loss, unspecified	L65.9
regnbuehinnebetendelse for 7 år siden. og ledgikt	Iridoicytites	H20
regnbuehinnebetendelse for 7 år siden. og ledgikt	Juvenile arthritis	M08
cyste he side i hodet, uspesifisert diagnose iflg pas	Sebaceous cyst	L72.3
Kreft i skjelbruskjertel. Op 9 mars 2010	Malignant neoplasm of thyroid gland	C73

¹ The original datapoint contain the name of this person. Redacted due privacy.

- a) Data access
- b) Data cleaning
- c) Questionary design

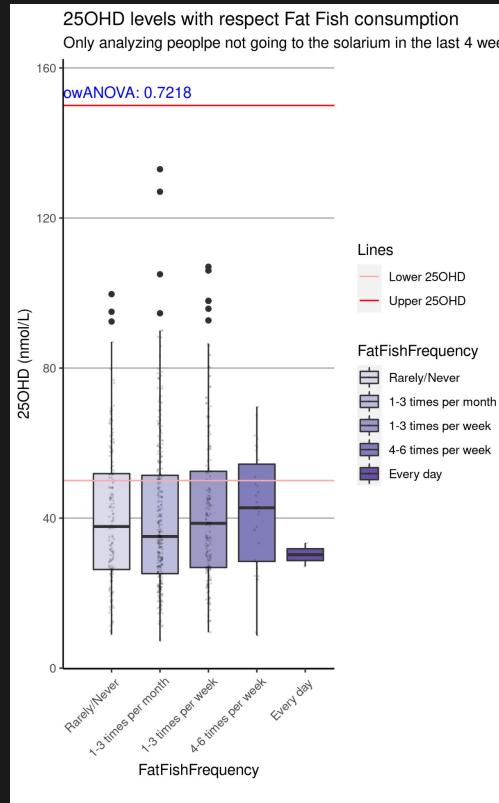
- a) Data access
- b) Data cleaning
- c) Questionary design

Sport activity is measure in METS (W/kg),
not in categorical variables!

	1	Never
	2	Less than daily
	3	1 per week
	4	2-3 per week
	5	4-6 per week
	6	Almost every day
	NA	Unknown
SportsFrequency	1	None
	2	About half an hour
	3	1 to 1.5 hours
	4	2 to 3 hours
	5	4 to 6 hours
	6	7 hours or more
	NA	Unknown
SportHours	1	Not hard at all
	2	A bit hard
	3	Quite hard
	4	Very hard
	5	Extremely hard
	NA	Unknown
SportsIntensity	1	By car or moped
	2	By bus
	3	By bike
	4	On foot
	NA	Unknown
SummerTransport		

- a) Data access
- b) Data cleaning
- c) Questionary design

Nutritional data has no validation and not proper 72h follow up.



- a) Data access
- b) Data cleaning
- c) Questionary design

BOYCE-CODD normal form!!

(3NF + Always satisfied lossless join conditions)

This is STATA and SPSS fault!



THE END
Thank you for comming!

-  Institutt for informatikk
-  rafael.a.canadas@uit.no
-  +47 77 64 56 85
-  Tromsø
-  REALF A 208

...questions?

