



Transportøkonomisk institutt
Stiftelsen Norsk senter for samferdselsforskning

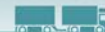
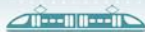


Hvor, og hvor mye sykles det i Oslo?

En estimering basert på imputering av sykkelteilerdata

Lars Even Egner, Christian Weber

1977/2023



Tittel:	Hvor, og hvor mye sykles det i Oslo? - En estimering basert på imputering av sykkelteilerdata
Tittel engelsk:	Where and how much cycling is done in Oslo? - An estimation based on imputation of bicycle counter data
Forfatter:	Lars Even Egner, Christian Weber
Dato:	09.2023
TØI-rapport:	1977/2023
Antall sider:	14
ISSN elektronisk:	2535-5104
ISBN elektronisk:	978-82-480-2037-0
Finansieringskilder:	Nordisk vegforum
TØIs p.nr.:	5357 – Sykkelimputering
Prosjektleder:	Lars Even Egner
Kvalitetsansvarlig:	Tor-Olav Nævestad
Fagfelt:	Sikkerhet og resiliens
Emneord:	Sykkel, reisevaneundersøkelser, sykkelteiler, app-data, imputering.

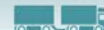
Kort sammendrag

Studien utforsker en ny metode for å estimere sykkelaktivitet i et område ved å kombinere data fra sykkelteiler og GPS-sykelstudier. Ved å modellere forholdet mellom disse datasettene og imputere manglende sykkelteilerdata, gir metoden absolutte estimater på hvor mange kilometer som sykles i Oslo, og hvor. Resultatene viser at syklister konsentrerer seg i Oslo sentrum, nordlige ringveier og E18 vestover. Gjennomsnittlig daglig sykling i området 5 km fra Oslo Domkirke er anslått til 1.115.913 km, omtrent 1,6 km per innbygger per dag. Selv om metoden er nyutviklet, er estimatene såpass realistiske at metoden med forsiktighet kan brukes i byplanlegging, skadeforebygging og folkehelseprosjekter.

Summary

The study explores a novel method to estimate bicycle activity in an area by combining data from bike counters and GPS bike studies. By modeling the relationship between these datasets and imputing missing bike count data, the method provides absolute estimates of the number of kilometers cycled in Oslo, and where. The results reveal that cyclists concentrate in downtown Oslo, northern ring roads, and E18 westward. The average daily cycling within 5 km of Oslo Cathedral is estimated at 1,115,913 km, roughly 1.6 km per resident per day. Despite being newly developed, the method's estimates are realistic enough to be cautiously applied in urban planning, injury prevention, and public health projects.

Transportøkonomisk institutt (TØI) har opphavsrett til hele rapporten og dens enkelte deler. Innholdet kan brukes som underlagsmateriale. Når rapporten siteres eller omtales, skal TØI oppgis som kilde med navn og rapportnummer. Rapporten kan ikke endres. Ved eventuell annen bruk må forhåndssamtykke fra TØI innhentes. For øvrig gjelder [Åndsverklovens](#) bestemmelser.



Forord

I denne rapporten presenterer vi en ny metode med medfølgende resultater på å estimere hvor mye, og hvor, det blir syklet i et område. Metoden avhenger av at det eksisterer både et tilstrekkelig antall sykkeltellere og GPS-sykkeldata i det samme området i samme periode. Ved å modellere sammenhengen mellom disse to datapunktene og imputere manglende sykkeltelldata på strekninger uten sykkeltellere, kan vi estimere hvor mye og hvor det blir syklet i dette området.

Lars Even Egner har vært hovedansvarlig på prosjektet, og stått for det meste av rapportskrivning og databehandling. Christian Weber har vært ansvarlig for innhenting av GIS-data, inklusiv underliggende veinett og passeringer på disse. Denne rapporten holdes forholdsvis kort, da prosjektets hovedformål er å utforske hvorvidt den foreslåtte metoden i det hele tatt er gjennomførbar. Vi takker Nordisk vegforums stipend 2023, som har finansiert prosjektet.

Oslo, september 2023
Transportøkonomisk institutt

Bjørne Grimsrud
Administrerende direktør

Trine Dale
Avdelingsleder



Innhold

Sammendrag

Summary

1	Innledning	1
2	Metode	2
	2.1 Datagrunnlag.....	2
	2.2 Databehandling.....	2
3	Resultater	4
4	Diskusjon og konklusjon	5
5	Referanser	6
	Vedlegg	7
	Vedlegg 1. Rstudio kode 1.....	7
	Vedlegg 2. Stata code	12
	Vedlegg 3. Rstudio kode 2.....	12

Hvor, og hvor mye sykles det i Oslo?

En estimering basert på imputering av sykkelteilerdata

TØI rapport 1977/2023 • Forfattere: Lars Even Egner, Christian Weber • Oslo 2023 • 14 sider

- Denne rapporten demonstrerer en ny metode for å estimere hvor, og hvor mye som blir syklet i et område.
- Metoden baserer seg på å modellere sammenhengen mellom antall sykler som passerer en veistrekning basert på sykkelteiler og eksisterende GPS-sykkeldata.
- Modellen estimerer at innen et 5km område fra Oslo Domkirke blir det i perioden 27-08-2018 til 01-10-2018 syklet 40,4 millioner kilometer, eller 1,6 kilometer per dag per innbygger i Oslo kommune.
- Selv om estimerer fra rapporten må brukes med noe forsiktighet da metoden er nyutviklet, viser denne rapporten at metoden er teknisk gjennomførbar for å estimere sykkelbruk i et område.

Where and how much cycling is done in Oslo?

An estimation based on imputation of bicycle counter data

TØI Report 1977/2023 • Authors: Lars Even Egner, Christian Weber • Oslo 2023 • 14 pages

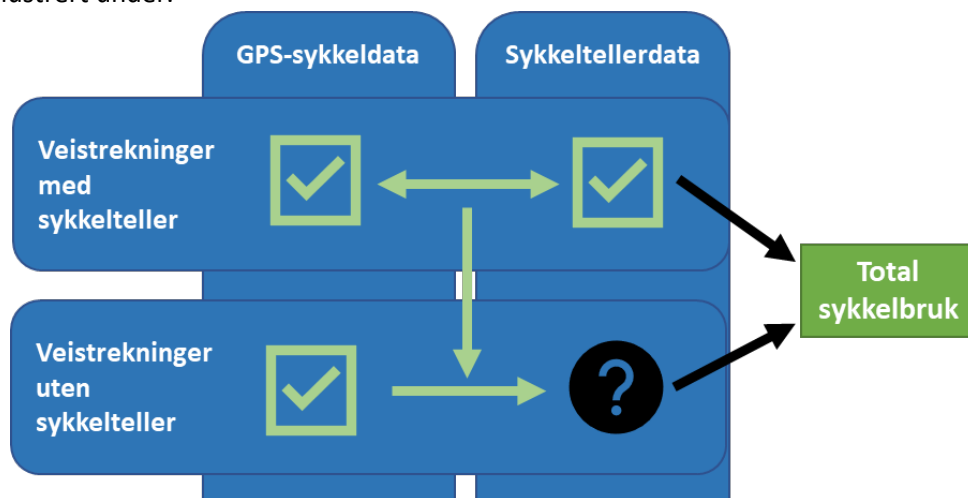
- This report demonstrates a new method for estimating where and how much cycling occurs in an area.
- The method is based on modelling the relationship between the number of bicycles passing a road segment between bike counters and existing GPS cycling data.
- The model estimates that within a 5km radius of Oslo Cathedral, during the period from 27-08-2018 to 01-10-2018, a total of 40.4 million kilometres were cycled, equivalent to 1.6 kilometres per day per inhabitant of the Oslo municipality.
- Although estimates from the report should be approached with caution due to the methods novelty, this report demonstrates the technical feasibility of using the method to estimate bicycle usage in an area.

1 Innledning

Nyere forskning har satt fokus på at sykkelskader er kraftig underrapportert og hvordan vi kan få bedre estimater på sykkelskader (Bjørnskau, 2021; Næss et al., 2020). Selv om vi nå nærmer oss bedre estimater på hvor mange som blir skadet, har vi ikke tilfredsstillende data på hvor mye det sykles. Næss et al. (2020) skriver at «It is difficult to attain data for bicycling exposure...». For å estimere hvor sikkert sykkel er som fremkomstmiddel, trenger man også å vite hvor mye det blir syklet.

Å si noe om hvor mye, og hvor det blir syklet, er metodisk utfordrende (Fyhri et al., 2022). Reisevaneundersøkelser kan fange opp relative endringer over tid, men denne vil alltid være i forhold til tidligere undersøkelser, og ikke absolutt. I tillegg sliter nye survey-baserte metoder med lave svarprosenten. Sykkeltellere kan si noe om hvor mange som reiser på en bestemt veibit, men data på enkelte veibiter gir et veldig begrensede svar på hvor og hvor mye det blir syklet totalt. GPS-studier, der deltakere lar seg spore med GPS, kan estimere relativt, men ikke absolutt bruk av veier. Alle metoder har styrker og svakheter, og alle har problemer med å estimere hvor mye det blir syklet totalt sett (Sørensen et al., 2015).

Ingen av metodene har til nå gitt gode resultater på absolutt sykkelbruk i et område, altså hvor mange kilometer det sykles i et område. Vi mener at imputering av sykkeltellerdata har potensiale for å gi dette. Imputering betyr å sette inn estimater for manglende verdier i datasett. For å imputere valide tall, trenger man et annet datasett å basere imputeringene på. Det er grunnlag for å tro at GPS-sykkeldata, antall passeringer på veistrekninger registrert i studier som anvender app-basert GPS data (e.g. de Jong et al., 2022) kan være dette grunnlaget. Sammenhengen imellom GPS-passeringer og sykkeltellerdata i veibiter med sykkeltellere, kan modelleres og estimeres i veibiter uten sykkeltellere. Den tenkte metoden er illustrert under.



Om metoden gir reliable resultater på hvor mye, og hvor, det sykles, kan den brukes i byplanlegging for å ta hensyn til mengden syklist i et område og planlegge sykkelinfrastruktur deretter. Med data over tid vil man kunne undersøke effekten av utbygging av sykkelinfrastruktur, både på rutevalg og hvor mye det sykles. Sykkeltrafikk kan også kombineres med skadedata for å generere nye estimater på risiko for sykkelulykker, dvs. hvor ofte sykkelskader skjer per kilometer som sykles. Om flere områder har både sykkeltellere og GPS-sykkeldata, vil man også kunne sammenligne sykkeladferd på tvers av byer. Store folkehelseprosjekter vil også kunne bruke sykkelbruk i et område som en kontrollerende variabel, eller undersøke effekten av dette direkte.

2 Metode

For å estimere sykkelbruk i Oslo-området, modellerte vi sammenhengen mellom sykkelpasseringer på alle veistrekninger med sykkeltellere i perioden 27-08-2018 til 01-10-2018, og sykkelpasseringer på samme veistrekning (fra de Jong et al., 2022). De Jong et al. (2022) samlet en database med 25.915 sykkelturner og vurderte i hvor stor grad syklistene bruker korteste vei. Det er dette datasettet vi også benytter oss av for å telle antall passeringer på veistrekninger. Med antakelsen om at deltakere fra studien ikke har nevneverdig forskjellig rutevalg fra alle andre syklistene, imputerte vi sykkelteilerdata på alle strekninger uten sykkelteiler basert på GPS data. Imputering vil si at vi oppretter «virtuelle» sykkelteiler på alle veistrekninger i området, og estimerer hva de hadde stått på basert på forholdet mellom ekte sykkelteiler og GPS-sykkelpasseringer.

2.1 Datagrunnlag

For å begrense området, arbeidet vi kun med veibiter som er mindre enn 5 kilometer fra Oslo domkirke. Dette området omfavner det meste av ring 3, og de fleste sykkelteiler i Oslo-området. Åpenbare båtlinjer i Oslofjorden ble ikke fjernet, dette for å inkludere syklistene som tar med sykkelen på båten.

Sykkeltellerdata ble innhentet fra Oslo kommunes sykkelteiler, og komplimentert med Statens Vegvesens sykkelteiler i det relevante området. Fordi sykkelteiler har vanskeligheter med å skille på syklistene og el-sparkeykeler (Fyhri et al., 2022), er det sannsynlig at estimatene i modellen er litt for høye. GPS-data ble innhentet fra en eksisterende studie (de Jong et al., 2022) som hadde kartlagt sykkelrutevalg blant 6.185 deltakere i Oslo. Sykkelerutenettverket er hentet fra samme studie. Dette nettverket basert på Open Street Map. Veisegmenter der sykling ikke er tillatt (eg. Motorveier), har fartsgrense over 60 km/t, eller har trapper, har blitt fjernet fra nettverket. I tillegg ble sterkt parallelle vei-segmenter slått sammen til et veisegment. Dette førte til at tre sykkelteiler som måler trafikk på en side av veien, der en annen måler en annen side av veien, ble slått sammen. Dette gjelder sykkelteiler Kirkevn ved Schonings gt. Sydgaende, Chr. Michelsens gt 53 Nord, og Ring 2 Kirkevn ved Sigyns gt. Sydgaende. Se de Jong (2022) for en nærmere beskrivelse av datainnhentingen

2.2 Databehandling

Sykkeltellerstasjoner ble manuelt koblet til det sykkelrutenettsegmentet som tilsvarte hvilke segment samme person i GPS og sykkelteilerdatasettet hadde blitt registrert på. Som regel var dette det nærmeste veisegmentet til sykkelteileren. Dette ga oss 1089 datapunkter. Vi gjennomførte en regresjonsmodell på sammenhengen mellom passeringer i sykkelteiler og GPS datasettet, fordelt på veistrekning og dato. Da sykkeldata er telle-data, valgte vi en Poisson regresjonsmodell. Dessverre ga denne modellen estimerer helt opp til 3.182065×10^{33} , som åpenbart er feil. Løsningen falt på en OLS regresjon uten konstantledd. Konstantleddet ble fjernet fordi det er en realistisk antagelse at ingen GPS passeringer vil også bety ingen andre syklistene, og det kan aldri være et negativt antall syklistene. Resultatene fra OLS regresjonen er presentert under.

Tabell 2.1: OLS regresjon av antall passeringer på sykkelteiler på veistrekke og dato. Justert $R^2 = 0.81$, $N = 1089$, model $p < .000$.

Variabel	Coef.	SE	p
Antall GPS-passeringer	62.3	1.98	< .000
Avstand fra Oslo domkirke i meter, invers	0.16	0.01	< .000

En forklart varians på 81% er tilstrekkelig for å imputere med relativt snevre estimater. Antall totale sykkelpasseringer pr. vegstrekning ble imputert ved å multiplisere antall GPS-passeringer og avstand fra Oslo domkirke med den respektive koeffisienten vist i tabell 1. Antall kilometer syklet er utregnet ved å multiplisere antall estimerte sykkelpasseringer med lengden på veistrekning. Databehandling ble gjennomført i R versjon 4.3.1 og Stata versjon 17. Det fulle statistiske skriptet er vedlagt i vedlegg 1 til 3.

3 Resultater

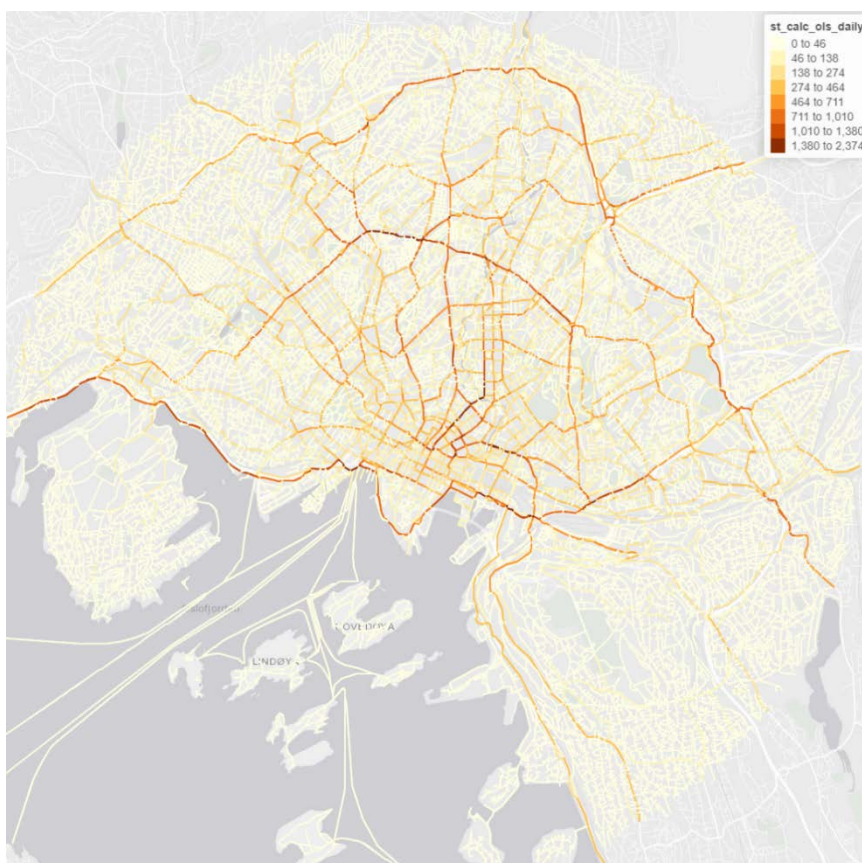
Tabell 2: Estimert antall sykkelkilometer tilbakelagt daglig i området.

Dato	Tusen kilometer syklet
27.08.2018	939
28.08.2018	1040
29.08.2018	1048
30.08.2018	1041
31.08.2018	1082
01.09.2018	964
02.09.2018	974
03.09.2018	1236
04.09.2018	1287
05.09.2018	1292
06.09.2018	1262
07.09.2018	1184
08.09.2018	991
09.09.2018	954
10.09.2018	1196
11.09.2018	1239
12.09.2018	1270
13.09.2018	1274
14.09.2018	1221
15.09.2018	1018
16.09.2018	972
17.09.2018	1254
18.09.2018	1282
19.09.2018	1290
20.09.2018	1272
21.09.2018	1115
22.09.2018	1012
23.09.2018	996
24.09.2018	1235
25.09.2018	1223
26.09.2018	1207
27.09.2018	1137
28.09.2018	1131
29.09.2018	956
30.09.2018	957
01.10.2018	895

Resultatene vises i tabell 2 som antall kilometer syklet i området pr. dag, og figur 1 som kart over hvor disse ble syklet. Kartet er også tilgjengelig på <https://www.toi.no/getfile.php?mmfileid=76132>.

Imputeringsmetoden viser konsentrasjoner av syklistene i Oslo sentrum ovenfor kvadraturen, nordre deler av ringveiene, og sykkelveien langs e18 vestover fra Oslo. Fordelt på dag, estimerer vi at det sykles i gjennomsnitt 1 115 913 kilometer per dag i området 5 kilometer fra Oslo Domkirke. Da Oslo kommune i 2018 hadde 673 469 innbyggere (SSB, 2023), tilsvarer dette 1.6 kilometer pr. person. Sammenlagt i perioden ble det tilbakelagt 40,4 millioner sykkelkilometer.

Estimeringsmetoden er nyutviklet og resultatene må derfor tolkes med noe forsiktighet.



Figur 3.1: Estimert gjennomsnittlig antall daglig sykkelpasseringer per veibit i perioden 27-08-2018 til 01-10-2018. "Sykling" i Oslofjorden representerer reiser der sykkel var hovedtransportmiddel.

4 Diskusjon og konklusjon

Denne studien viser at å imputere sykkelteilerdata på grunnlag av GPS-data kan være en gyldig måte å få frem hvor mye som blir syklet i et område, målt i antall kilometer sykling. Kombinert med komplette skadedata for syklistere (se Bjørnskau, 2021), skal metoden i utgangspunktet være mulig å bruke for å estimere hvor ofte ulykker skjer på sykkel. Funnene er sammenlignbare med for eksempel reisevaneundersøkelsen i Oslo, som viser at en gjennomsnittlig Osloborger har 0,22 sykkelreiser daglig, der disse er 4,6km lange (Ellis et al., 2022). Dette gir 1km syklet per person per dag hele året. Siden man kan anta det sykles mindre på vinteren, er dette sammenlignbart med våre funn på 1,6 kilometer per person. Merk også at vårt estimat reduseres av at vi inkluderer alle borgere, imens reisevaneundersøkelsen kun inkluderer innbyggere over 13 år. I tillegg er områdene (2km fra Oslo domkirke og Oslo kommune), ikke er helt sammenlignbare. Oppfølgingsstudier vil kunne også se på utvikling i sykkelbruk med nye metoder.

Metoden kan videreutvikles ved å inkludere andre variabler som estimerer sykkelbruk. Disse inkluderer el-sparkesykkeldata, sentralitet til veistrekning, værdata, elevasjon på veistrekningen og mer. Den beste måten å forbedre estimatene til metoden på, vil sannsynligvis være å utvide innsamling av sykkelteilerdata i områder det blir syklet lite. Dette må sannsynligvis gjøres ved manuelle tellinger. Ved innsamling av GPS sykkeldata for lengre perioder, kan metoden brukes til å estimere sykling i lengre perioder. For nå kan metoden kun gi estimater for den gitte perioden vi hadde GPS sykkeldata tilgjengelig. Metoden kan utvikles til å gi estimater for mer presise områder enn 5 kilometer fra Oslo domkirke. Grunnkretser, dekningsområder for legevakter, og mer kan være relevante. De eneste kravene er at det eksisterer GPS-sykkeldata og tilstrekkelig antall sykkelteiler i området.

Oppsummert vil vi si at imputering av sykkelteilerdata basert på GPS-sykkeldata virker å være en realistisk måte å få fram totalt antall kilometer det sykles i et område. Innenfor et område på 5 kilometer fra Oslo domkirke, ble det i perioden 27-08-2018 til 01-10-2018, en god sykkelperiode, syklet 40,4 millioner kilometer. Selv om metoden, så langt vi vet, er nyutviklet, virker tallene den produserer realistiske. Metoden kan med fordel utvikles med mer data og flere indikatorer, men dette vil kreve mer ressurser og videre forskning.

5 Referanser

- Bjørnskau, T. (2021). *Trafikksikkerhet for syklister og fotgjengere – status og utfordringer* (No. 1844/2021). Transportøkonomisk Institutt.
- de Jong, T., Böcker, L., & Weber, C. (2022). Road infrastructures, spatial surroundings, and the demand and route choices for cycling: Evidence from a GPS-based mode detection study from Oslo, Norway. *Environment and Planning. B, Urban Analytics and City Science*, 239980832211414.
- Ellis, I. O., Strætkvern, A., Berglund, G., & Kjørstad, K. (2022). *Reisevaner i Oslo og Viken. En analyse av nasjonal reisevaneundersøkelse 2018/19*. PROSAM.
- Fyhri, A., Ellis, I. O., Pokorny, P., de Jong, T., & Weber, C. (2022). *Nå telte han deg óg! Hvordan måle sykling og nye former for mikromobilitet* (No. 1897). Transportøkonomisk institutt.
- Næss, I., Galteland, P., Skaga, N. O., Eken, T., Helseth, E., & Ramm-Pettersen, J. (2020). The number of patients hospitalized with bicycle injuries is increasing - A cry for better road safety. *Accident; Analysis and Prevention*, 148(105836), 105836.
- Sørensen, M. W. J., Bjørnskau, T., Fyhri, A., & de Jong, T. (2015). *Før-og etterundersøkelser av sykkeltiltak* (No. 1392/2015). Transportøkonomisk Institutt.
<https://www.toi.no/getfile.php?mmfileid=40079>
- SSB. (2023, February 21). *Befolkning*. Statistisk Sentralbyrå. <https://www.ssb.no/statbank/table/07459>

Vedlegg

Vedlegg 1. Rstudio kode 1

```

library(tidyverse)
library(sf)
library(terra)
library(tmap)
Sys.setlocale(locale = "no_NO")
tmap_mode("view")

bike_counter_path <- "I:/SM-AVD/5357 Sykkelimputering/data/export.csv"
gis_data_path <- "I:/TEK-
AVD/Forskermapper/CWE/python_GIT/analysis_senseDat/5158Sykkeltelling/o
utput/senseDat_Passings_2018.geojson"
output_path <- "I:/SM-AVD/5357 Sykkelimputering/data/combo_data.csv"

### Bike counters #####
# Import SVV data
setwd("I:/SM-AVD/5357 Sykkelimputering/data/SVV data")
SVV_BK <- read_csv2
csv_files <- list.files(pattern = "\\*.csv$")
data_list <- lapply(csv_files, function(file) read_csv2(file, locale =
locale(encoding = "UTF-8")))

SSV_ST_all <- do.call(rbind, data_list)
SSV_ST_all2 <- SSV_ST_all %>%
  filter(Felt == "Totalt")

# Rename to match other datasets and delete unused variables.
SSV_ST_all3 <- SSV_ST_all2 %>%
  rename(streetName = Navn) %>%
  rename(startDate = Dato) %>%
  rename(STdata = Trafikkmengde) %>%
  select(streetName, startDate, STdata)

# Rename, obsolete
SSV_ST_all4 <- SSV_ST_all3

# Oslo dataset
# Import
bike_counters <- read_csv2(bike_counter_path,
  skip = 2,
  col_names = TRUE
)

# Transpose the dataset
bike_counters_flipped <- t(bike_counters)
colnames(bike_counters_flipped) <-
as.character(bike_counters_flipped[1, ]) # Set the first row as column
names

```

Hvor, og hvor mye sykles det i Oslo?

```
bike_counters_flipped <- bike_counters_flipped[-1, ] # Remove the
first row (since it was used for column names)
bike_counters_flipped <- as.data.frame(bike_counters_flipped) # Change
to data frame.
bike_counters_flipped <- rownames_to_column(bike_counters_flipped, var
= "streetName") #Change the row names to a column/variable.

#Pivot longer
bike_counters_flipped <- bike_counters_flipped %>%
  pivot_longer(cols = starts_with("20"), names_to = "startDate",
values_to = "STdata") %>%
  mutate(startDate = as.Date(startDate))

# Delete some observations that are not needed.
bike_counters_flipped <- bike_counters_flipped %>%
  filter(!is.na(STdata)) %>% # Delete observations with NA.
  filter(streetName != "Alle sykkelteellere a")

#Convert to numeric
bike_counters_flipped$STdata <-
as.numeric(bike_counters_flipped$STdata)

#Merge
ST_data_all <- bind_rows(SSV_ST_all4, bike_counters_flipped)

#Convert from øå til aa o aa
ST_data_all <- ST_data_all %>%
  mutate(streetName = str_replace_all(streetName, "\u00e5", "aa")) %>%
  mutate(streetName = str_replace_all(streetName, "\u00f8", "o")) %>%
  mutate(streetName = str_replace_all(streetName, "\u00e6", "ae")) %>%
  mutate(streetName = str_replace_all(streetName, "\u00c5", "aa")) %>%
  mutate(streetName = str_replace_all(streetName, "\u00d8", "o")) %>%
  mutate(streetName = str_replace_all(streetName, "\u00c6", "ae")) %>%
  mutate(streetName = str_replace_all(streetName, "[\u00e5\u00e5]",
"aa")) %>%
  mutate(streetName = str_replace_all(streetName, "[\u00f8\u00d8]",
"o")) %>%
  mutate(streetName = str_replace_all(streetName, "[\u00e6\u00c6]",
"ae")) %>%
  mutate(streetName = str_replace_all(streetName, "[\u00e5]", "ae")) %>%
  mutate(streetName = str_replace_all(streetName, "[\u00f8]", "o")) %>%
  mutate(streetName = str_replace_all(streetName, "[\u00c5]", "aa")) %>%
  mutate(streetName = str_replace_all(streetName, "\u00E5", "aa")) %>%
  mutate(streetName = str_replace_all(streetName, "\u00F8", "o")) %>%
  mutate(streetName = str_replace_all(streetName, "\u00E6", "ae")) %>%
  mutate(streetName = str_replace_all(streetName, "\u00C5", "aa")) %>%
  mutate(streetName = str_replace_all(streetName, "\u00D8", "o")) %>%
  mutate(streetName = str_replace_all(streetName, "\u00C6", "ae")) %>%
  mutate(streetName = str_replace_all(streetName, "\u2013", "aa"))

### GIS data.#####
# Import
gis_data <- st_read(gis_data_path)
```



```

st_crs(gis_data) <- 32632 #Data is stored in wrong UTM. Changing to
uTM32.

# Remove everything further away than 5km from Oslo domkirke
lon <- 10.7464338447269
lat <- 59.91278076251841
point_data <- data.frame(lon = lon, lat = lat)
point_sf <- st_as_sf(point_data, coords = c("lon", "lat"), crs = 4326)
%>%
  st_transform(32632)# Transform to SF
gis_data$distances <- st_distance(gis_data, point_sf) # Calculate
distances from each linestring to the location
gis_data$distances <- as.numeric(gsub("\\s\\[m\\]", "",
gis_data$distances)) # Remove "[m]" and convert to numeric
gis_data_2 <- filter(gis_data, gis_data$distances <= 5000) # Remove
everything over 5k

# Store the distance in a separate dataset
gis_data_2_dist <- gis_data_2 %>%
  select(id, distances) %>%
  st_drop_geometry() %>%
  distinct()

# Generate a map only object to match counters to road ID.
duplicates <- duplicated(gis_data_2$id)
gis_data_2MO <- gis_data_2[!duplicates, ]
gis_data_2MO <- gis_data_2MO %>%
  select(-startDate) %>%
  select(-MMAbs) %>%
  select(-distances) %>%
  select(-id_orig)

# Generate the length of each line in a separate dataset, from the map
only sf object.
gis_data_2LL <- gis_data_2MO %>%
  mutate(line_length = st_length(.)) %>%
  st_drop_geometry()
gis_data_2LL$line_length <- as.numeric(gsub("\\s\\[m\\]", "",
gis_data_2LL$line_length)) # Remove "[m]" and convert to numeric

# Remove the GIS parts because they are no longer needed. Also remove
all non needed things.
gis_data_2.1 <- gis_data_2 %>%
  st_drop_geometry() %>%
  select(-id_orig) %>%
  select(-distances)

# Change all NA GISdata to 0.
gis_data_2.2 <- gis_data_2.1 %>%
  mutate(MMAbs = ifelse(is.na(MMAbs), 0, MMAbs))

# Generate a dataset with every date x id combination, and join them.
dates_df <- data.frame(startDate = seq.Date(as.Date("2018-08-27"),
as.Date("2018-10-01"), by = "day"))
dates_x_ID <- gis_data_2.2 %>%

```

Hvor, og hvor mye sykles det i Oslo?

```
distinct(id, .keep_all = TRUE) %>%
select(id) %>%
cross_join(dates_df) %>%
arrange(id, startDate)
gis_data_3 <- dates_x_ID %>%
left_join(gis_data_2.2, by = c("id", "startDate")) %>%
mutate(MMAbs = ifelse(is.na(MMAbs), 0, MMAbs))

#Rename to more reasonable things and add a length variable
gis_data_4 <- gis_data_3 %>%
rename(GPSdata = MMAbs)

# Merge the line-length dataset and the observations pr. date and id.
gis_data_5 <- merge(gis_data_4, gis_data_2LL, by = "id", all.x =
FALSE)

##### Attach a ID's to the bike dataset. #####
ST_data_all2 <- ST_data_all %>%
mutate(id = case_when(
streetName == "Bislettgata 1" ~ 37962,
streetName == "Bygdoy alle 13" ~ 58710,
streetName == "Bispegata" ~ 40173,
streetName == "Baerumsveien 22" ~ 0,
streetName == "Chr Michelsens gt sydside" ~ 11882,
streetName == "Chr. Michelsens gt 53 Nord" ~ 11882,
streetName == "Ekebergveien 153" ~ 276895,
streetName == "Eventyrbrua" ~ 28011,
streetName == "Griffenfeldts gate 19" ~ 19857,
streetName == "Hellerudveien 76 v/ trasop skole" ~ 0,
streetName == "Hoffsveien 40" ~ 81798,
streetName == "Holmenkollvn 42" ~ 0,
streetName == "Jon Smestads vei 2" ~ 273674,
streetName == "Kierschowsgate 10" ~ 24754,
streetName == "Kongsveien 48" ~ 50707,
streetName == "Maridalsveien 16 ved Vulkan" ~ 227715,
streetName == "Maridalsveien 323 ved Korsvollbakken" ~ 11830,
streetName == "Monolittveien ved Frognerparken" ~ 272738,
streetName == "Nordstrandveien 59" ~ 0,
streetName == "Ring 2 Kirkevn ved Schonings gt Nordgaaende" ~
46754,
streetName == "Ring 2 Kirkevn ved Schonings gt. Sydgaaende" ~
46754,
streetName == "Ring 2 Kirkevn ved Sigyns gt. Nordgaaende" ~ 68041,
streetName == "Ring 2 Kirkevn ved Sigyns gt. Sydgaaende" ~ 68041,
streetName == "Raadhusgata 28" ~ 46799,
streetName == "Thorvald Meyers gt. 11" ~ 18811,
streetName == "Tvetenveien 2 gang og sykkelsti" ~ 20099,
streetName == "Taasenveien 43" ~ 224554,
streetName == "Ullern gaardsvei 41" ~ 0,
streetName == "Ullevaalsveien 8" ~ 234721,
streetName == "Vaterlands bro" ~ 31762,
streetName == "Vaekeroveien 148" ~ 0,
streetName == "Waldemar Thranes gt 51" ~ 25523,
streetName == "Wergelandsveien" ~ 44381,
```

```

streetName == "aakebergveien 28" ~ 229358,
streetName == "ostensjoveien v/Brynseng st" ~ 22879,
streetName == "TELLER X3 - Vei og Sykkelfelt retning mot Storo" ~
13095, ## NOTE: Has some zero counts that seem off.
streetName == "ostensjovn ved ostensjovannet" ~ 0,
streetName == "Smestad Sykkel" ~ 260325,
streetName == "Dr. Eufemias gt. Sykkel Vestgaaende" ~ 40166,
streetName == "Munkedamsveien Sykkel" ~ 54600,
streetName == "Bryn Sykkel" ~ 228408,
streetName == "AKERSYKE. SYKKEL 03" ~ 3228,
streetName == "FROGNERST.SYKKEL 03" ~ 64662,
streetName == "Dr. Eufemias Gt. Sykkel aastgaaende" ~ 40167,
streetName == "Gaustad sykkel" ~ 37883, ## Note: Counter has been
somewhat moved, as the road network does not match the actual road.
Moved so that there should be a high correlation between people moving
through the bike counter and this road segment.
# Note: Bike counter "Tåsen" is nowhere to be found, Possible
third dataset with more counters?
TRUE ~ NA_integer_
))

# Delete all observations with ID == 0, meaning its outside our
intended range, and replace some 0 counts in bike counters with NA.
ST_data_all3 <- ST_data_all2 %>%
  filter(id != 0) %>%
  mutate(STdata = ifelse(STdata == 0, NA, STdata))

# Combine STdata of duplicate ID's. Delete some specific observations
with counters two ways.
ST_data_all4 <- ST_data_all3 %>%
  group_by(id, startDate) %>%
  mutate(STdata = sum(STdata)) %>%
  filter(streetName != "Ring 2 Kirkevn ved Schonings gt. Sydgaende")
%>%
  filter(streetName != "Chr. Michelsens gt 53 Nord") %>%
  filter(streetName != "Ring 2 Kirkevn ved Sigyns gt. Sydgaende") %>%
  filter(streetName != "TELLER X6") %>% ## Not a duplicate, but only
have 2 observations with 0 passings so yeah.
  filter(streetName != "Ullevaalsveien 8") %>% ## Not a duplicate, but
only have 0-count observations, which seem unrealistic for such a
central location.
  filter(streetName != "aakebergveien 28") ## Not a duplicate, but
have way to low counts to be real. Most are 0, rest are between 1 and
4.

# Combine the large GIS dataset with the small bike counter dataset.
combo_data <- left_join(gis_data_5, ST_data_all4, by = c("id",
"startDate"))

# Attach the distance again
combo_data2 <- full_join(combo_data, gis_data_2_dist, by = "id")

# Export as a CSV
write.csv(combo_data2, file = output_path, row.names = FALSE)

```

Hvor, og hvor mye sykles det i Oslo?

Vedlegg 2. Stata code

```
version 17
clear all
import delimited "I:\SM-AVD\5357
Sykkelimputering\data\combo_data.csv", numericcols(6)

// Format to date and rename
generate date = date(startdate, "YMD")
format date %td
rename stdata STdata
rename gpsdata GPSdata
rename startdate startDate

// Reverse distance, as further away from Oslo should mean less
traffic
gen dist_rev = 5000 - distances

//Check connection between variables.
poisson STdata GPSdata
poisson STdata GPSdata dist_rev
poisson STdata GPSdata, noconstant
poisson STdata GPSdata dist_rev, noconstant

// Even though Poisson makes more theoretical sense, it creates some
WILD estimates for the upper end. OLS estimates are way more
reasonable, especially the noconstant ones.
reg STdata GPSdata dist_rev
reg STdata GPSdata dist_rev, noconstant

//Impute, if we in the future want to show uncertainty with varying
datasets.
mi set wide
mi register impute STdata
mi register regular GPSdata distances
mi register passive id line_length date
mi impute monotone (poisson, noconstant) STdata = GPSdata distances,
add(5) dots

// Export
drop date
export delimited using "I:\SM-AVD\5357
Sykkelimputering\data\imputed_dataset.csv", replace
```

Vedlegg 3. Rstudio kode 2

```
library(tidyverse)
library(sf)
library(terra)
library(tmap)
library(RColorBrewer)

tmap_mode("view")
```

```

imputed_dataset_path <- "I:/SM-AVD/5357
Sykkelimputering/data/imputed_dataset.csv"
gis_data_path <- "I:/TEK-
AVD/Forskermapper/CWE/python_GIT/analysis_senseDat/5158Sykkeltelling/o
utput/senseDat_Passings_2018.geojson"

##### GIS DATA #####
#####
gis_data <- st_read(gis_data_path)
st_crs(gis_data) <- 32632 #Data is stored in wrong UTM. Changing to
uTM32.
# Remove everything further away than 5km from Oslo domkirke
lon <- 10.7464338447269
lat <- 59.91278076251841
point_data <- data.frame(lon = lon, lat = lat)
point_sf <- st_as_sf(point_data, coords = c("lon", "lat"), crs = 4326)
%>%
  st_transform(32632)# Transform to SF
gis_data$distances <- st_distance(gis_data, point_sf) # Calculate
distances from each linestring to the location
gis_data$distances <- as.numeric(gsub("\\s\\[m\\]", "",
gis_data$distances)) # Remove "[m]" and convert to numeric
gis_data_2 <- filter(gis_data, gis_data$distances <= 5000) # Remove
everything over 5k

# Remove wrong ID, dates and distance
gis_data_2.1 <- gis_data_2 %>%
  select(-distances) %>%
  select(-id_orig) %>%
  select(-startDate)

# Change all NA GISdata to 0. Rename some.
gis_data_2.2 <- gis_data_2.1 %>%
  mutate(MMAbs = ifelse(is.na(MMAbs), 0, MMAbs)) %>%
  rename(GPSdata = MMAbs)

# Remove duplicates
duplicates <- duplicated(gis_data_2.2$id)
gis_data_3 <- gis_data_2.2[!duplicates, ]

##### imputed dataset work #####
#####
imp_set <- read_csv(imputed_dataset_path)

# Potentially unalive some lines on the fjord, wait and see. Original
scientific article keeps them, so we do too.
imp_set1.5 <- imp_set

# Create a separate branch dataset based on days not period
daily_bike_km <- imp_set1.5 %>%
  mutate(st_calc_ols = (GPSdata * 62.31642) + ((5000 - distances) *
0.1647417)) %>%
  group_by(startDate) %>%

```

Hvor, og hvor mye sykles det i Oslo?

```
summarise(km_trav_day = sum(st_calc_ols * line_length) / 1000)

# Create a summary variable for passings on road in period.
imp_set2 <- imp_set1.5 %>%
  group_by(id) %>%
  mutate(sum_imp1_STdata = sum(`_1_STdata`, na.rm = TRUE)) %>%
  mutate(sum_GPSdata = sum(`GPSdata`, na.rm = TRUE)) %>%
  mutate(sum_STdata = sum(`STdata`, na.rm = TRUE))

# Remove all variables we dont need.
imp_set3 <- imp_set2[, c("id", "sum_imp1_STdata", "sum_GPSdata",
"sum_STdata", "distances")]

# Remove all duplicates
duplicates <- duplicated(imp_set3$id)
imp_set4 <- imp_set3[!duplicates, ]

# Join the two datasets.
full_set <- right_join(gis_data_3, imp_set4, by = c("id"))

# Calculate estimated based on regression models.
full_set2 <- full_set %>%
  mutate(st_calc_poi = exp(sum_GPSdata * 0.0565149) + exp((5000 -
distances) * 0.001616)) %>%
  mutate(st_calc_ols = (sum_GPSdata * 62.31642) + ((5000 - distances)
* 0.1647417))

# Generate a daily number.
# First calculate the number of days in the dataset.
date_range <- imp_set %>%
  summarise(min_date = min(startDate),
            max_date = max(startDate)) %>%
  mutate(number_of_days = as.numeric(difftime(max_date, min_date,
units = "days")) + 1) %>%
  pull(number_of_days)

# Then divide by that number.
full_set3 <- full_set2 %>%
  mutate(st_calc_poi_daily = round(st_calc_poi / date_range)) %>%
  mutate(st_calc_ols_daily = round(st_calc_ols / date_range))

# Define the number of breaks
num_breaks <- 8

# Get a color palette suitable for lines on a bright background
color_palette <- brewer.pal(n = num_breaks, name = "YlOrBr")

# Create the thematic map with natural breaks
tm_shape(full_set3) +
  tm_lines(col = "st_calc_ols_daily", lwd = 2,
           palette = color_palette, style = "fisher", n = num_breaks)
+
  tm_layout(frame = FALSE)
```


TØI er et anvendt forskningsinstitutt som mottar basisbevilgning fra Norges forskningsråd og gjennomfører forsknings- og utredningsoppdrag for næringsliv og offentlige etater. TØI ble opprettet i 1964 og er organisert som uavhengig stiftelse.

TØI utvikler og formidler kunnskap om samferdsel med vitenskapelig kvalitet og praktisk anvendelse. Instituttet har et tverrfaglig miljø med rundt 90 høyt spesialiserte forskere.

Instituttet driver forskningsformidling gjennom TØI-rapporter, artikler i vitenskapelige tidsskrifter, bøker, seminarer, samt innlegg og intervjuer i media. TØI-rapportene er gratis tilgjengelige på instituttets hjemmeside www.toi.no.

TØI er partner i CIENS Forskningscenter for miljø og samfunn, lokalisert i Forskningsparken nær Universitetet i Oslo (se www.ciens.no). Instituttet deltar aktivt i internasjonalt forskningssamarbeid, med særlig vekt på EUs rammeprogrammer.

TØI dekker alle transportmidler og temaområder innen samferdsel, inkludert trafiksikkerhet, kollektivtransport, klima og miljø, reiseliv, reisevaner og reiseetterspørsel, arealplanlegging, ITS, offentlige beslutningsprosesser, næringslivets transportbehov og generell transportøkonomi.

Transportøkonomisk institutt krever opphavsrett til egne arbeider og legger vekt på å opptre uavhengig av oppdragsgiverne i alle faglige analyser og vurderinger.

Postadresse:

Transportøkonomisk institutt
Gautstadalléen 21
0349 Oslo
Norge

E-post: toi@toi.no

Kontoradresse:

Forskningsparken
Gautstadalléen 21

Hjemmeside: www.toi.no

