MSB106 - Assignment

Sindre H. Øveraas, Alen Colakovic, Mona Lisa Jones & Sebastian M. Fløysand

```
library(rgdal)
library(dplyr)
library(RSQLite)
library(sf)
library(tidyverse)
library(readr)
library(ggplot2)
library(kableExtra)
library(huxtable)
```

```
NOR_CBD <- read_csv("NOR_CBD.csv")
Dist_CBD_Dentist <- read_csv("Dist_CBD_Dentist.csv")
```

```
Warning: One or more parsing issues, see `problems()` for details
```

```
Zonal_CBD <- read_csv("Zonal_CBD.csv")
Dist_Mal_Dentist <- read.csv("Dist_Mal_Dentist.csv")
NOR_OSM_SHOP_MAL <- read.csv("NOR_OSM_SHOP_MAL.csv")
Zonal_Den <- read.csv("Zonal_Den.csv") |>
  select(fid, X_sum)
NOR_KOMM <- read_csv("NOR_KOMM.csv")
NOR_KOMM_2 <- read_csv("NOR_KOMM_2.csv")
```

```
Commune_Data <- inner_join(NOR_KOMM, NOR_KOMM_2, by = "kommunenummer") |>
  select(kommunenummer, HubName, HubDist, Turnover_capita_retail_Omsetning) |>
  rename("knr" = kommunenummer, "DistMal" = HubDist, "HubNameMal" = HubName, "Turnover_capita_retail
```

```
Dentist_Data <- inner_join(Dist_Mal_Dentist, Dist_CBD_Dentist, by = "fid") |>
  select(fid, Juridisk.n, Antall.ans, Sum.Drifts, Sum.salgsi, Driftsresu.y, osm_id.y, latlong.y, Hubl
  rename("DistMal" = HubDist.x, "DistCBD" = HubDist.y, "HubNameMal" = HubName.x, "HubNameCBD" = HubNa
```

```
Dentist_Data <- inner_join(Dentist_Data, Zonal_Den, by = "fid")
```

Introduction

Our main task in this assignment is to perform an analysis of geospatial determinants of firm activity. More specifically we are to focus on the Norwegian dental industry in this regard, and see how geospatial determinants such as distances to shopping malls and CBDs (Central Business Districts), as well as population density can determine dental businesses income and general financial operations. As an example, central questions in this assignment will be; "Is it more beneficial to be highly centralized in urban areas with high population density and many competitors, or is it a greater advantage to be less centralized to the advantage that the nearest competing company is considerably further away?", "Which determinants appear to be most significant for economic benefit?"

Theoretical Foundation of Hypothesis

Data description

```
kable(summary(Commune_Data))
```

	knr	HubNameMal	DistMal	Turnover_capita_retail
	Length:435	Min. : 6.0	Min. : 0.01854	Min. : 17366
	Class :character	1st Qu.:220.0	1st Qu.: 9.41403	1st Qu.: 59250
	Mode :character	Median :291.0	Median :18.87373	Median : 83203
	NA	Mean :280.8	Mean :23.82982	Mean : 84719
	NA	3rd Qu.:363.0	3rd Qu.:32.67127	3rd Qu.:106257
	NA	Max. :424.0	Max. :98.37355	Max. :218728

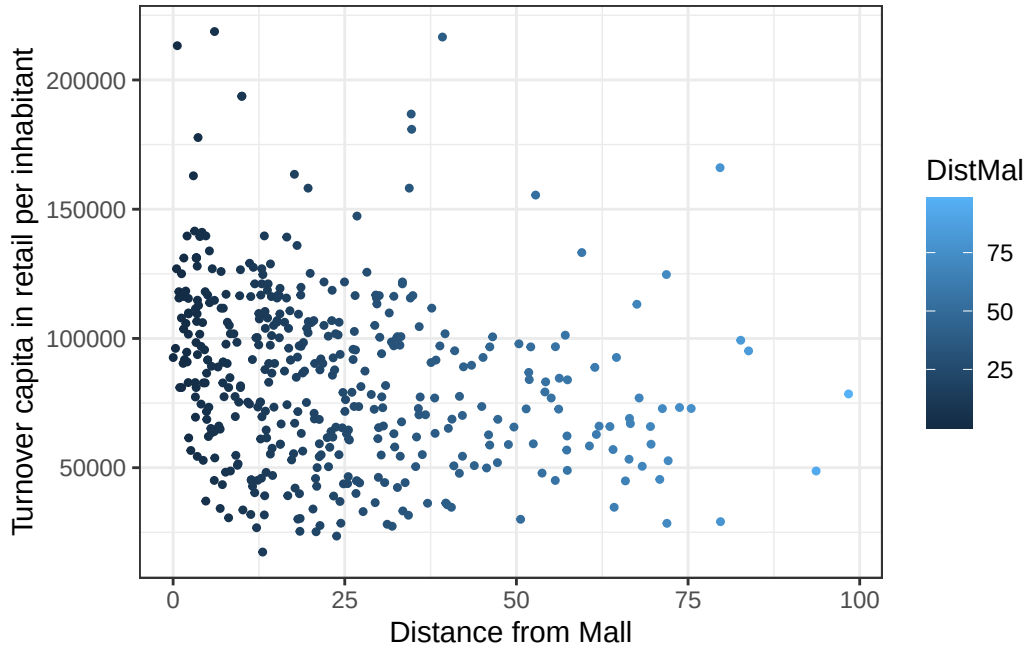
```
kable(summary(Dentist_Data[3:6]))
```

	Antall.ans	Sum.Drifts	Sum.salgsi	Driftsres
	Min. : 0.00	Min. :0.000e+00	Min. : -28000	Min. :-30319000
	1st Qu.: 1.00	1st Qu.:1.117e+06	1st Qu.: 658250	1st Qu.: -5000
	Median : 1.00	Median :3.616e+06	Median : 3176500	Median : 330000
	Mean : 39.26	Mean :5.085e+07	Mean : 50123495	Mean : 4842620
	3rd Qu.: 3.00	3rd Qu.:6.820e+06	3rd Qu.: 6248250	3rd Qu.: 1201750
	Max. :6870.00	Max. :1.058e+09	Max. :1058102000	Max. :108100000
	NA	NA's :3838	NA's :3838	NA's :3838

```
kable(summary(Dentist_Data[8:12]))
```

	latlong	HubNameCBD	DistCBD	HubNameMal	DistMal
	Length:5740	Min. : 1.00	Min. : 0.0046	Min. : 3.0	Min. : 0.0008
	Class :character	1st Qu.:14.00	1st Qu.: 1.3912	1st Qu.:130.0	1st Qu.: 0.3145
	Mode :character	Median :19.00	Median : 6.7067	Median :211.0	Median : 0.9164
	NA	Mean :17.14	Mean : 20.5000	Mean :214.5	Mean : 4.6331
	NA	3rd Qu.:20.00	3rd Qu.: 19.8885	3rd Qu.:291.0	3rd Qu.: 2.5220
	NA	Max. :33.00	Max. :470.1005	Max. :424.0	Max. :105.8867

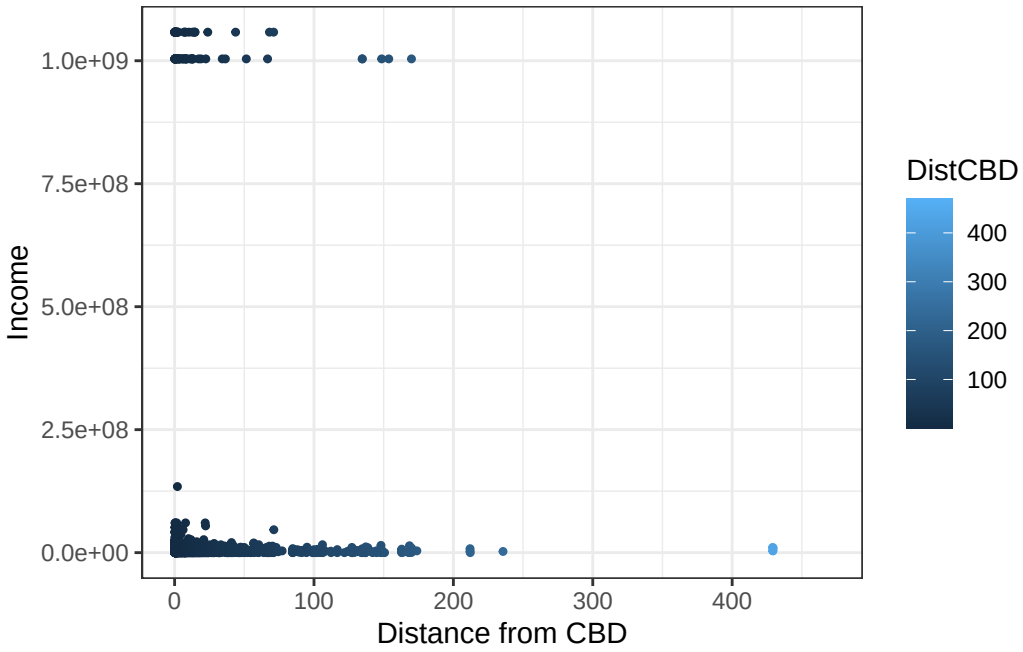
```
Commune_Data |>
  ggplot(aes(x = DistMal, y = Turnover_capita_retail, colour = DistMal)) +
  geom_point(lwd = .9) +
  labs(x = "Distance from Mall", y = "Turnover capita in retail per inhabitant") +
  theme_bw()
```



Might be more appropriate to explore the distance from CBDs to Malls compared to turnover capita in retail per inhabitant, rather than the distance from the geographical centroid of Norwegian Communes, because of the tendency of bigger population density around CBDs.

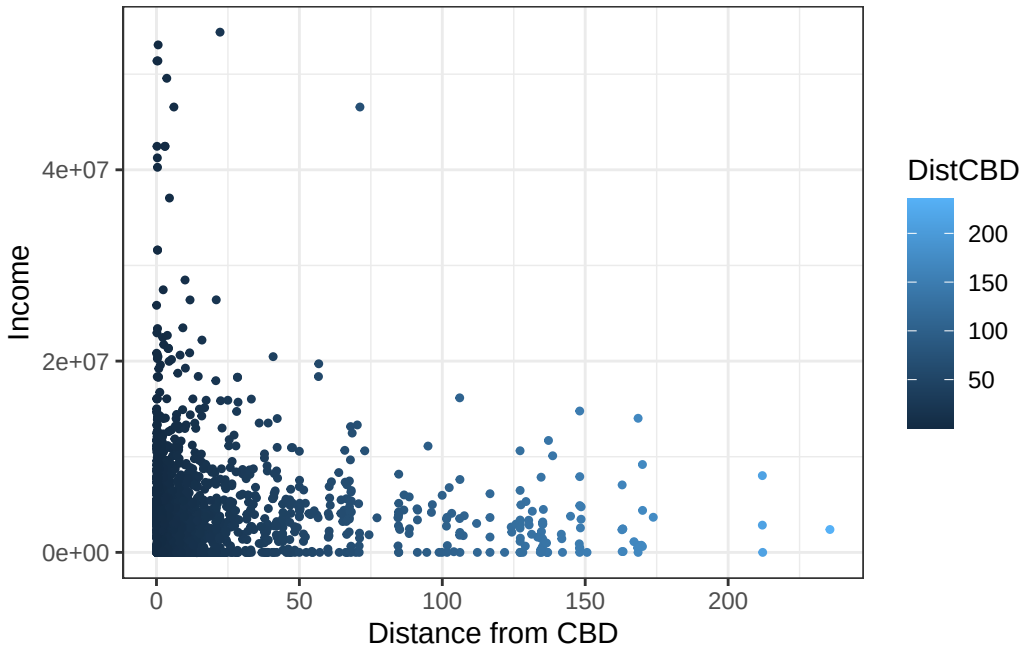
```
Dentist_Data |>
  ggplot(aes(x = DistCBD, y = Sum.salgsi, colour = DistCBD)) +
  geom_point(lwd = .9) +
  labs(x = "Distance from CBD", y = "Income") +
  theme_bw()
```

```
Warning: Removed 3838 rows containing missing values (geom_point).
```



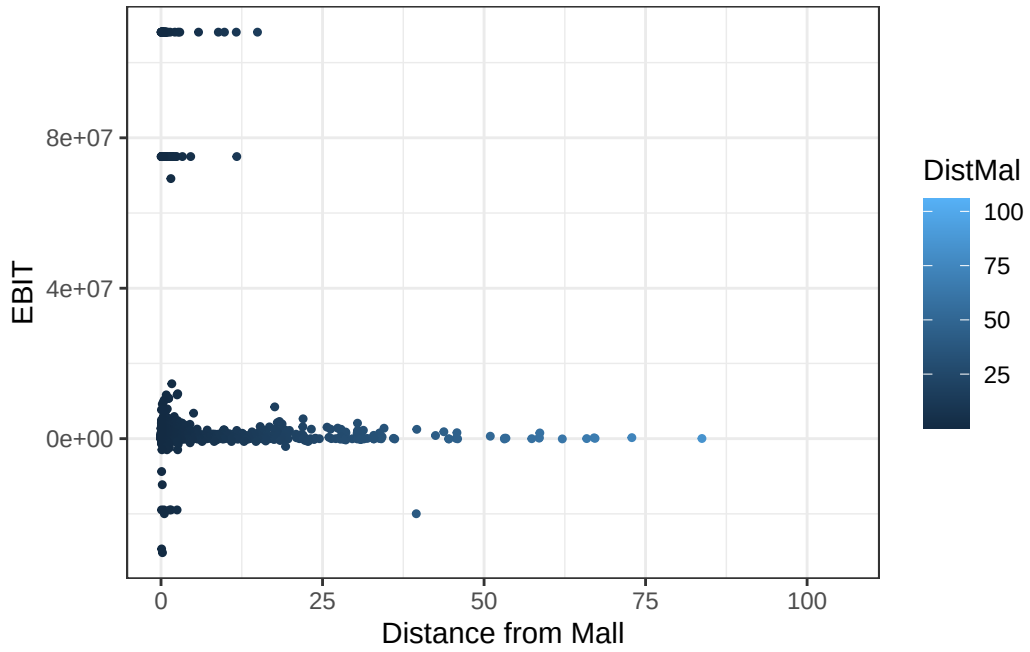

```
Dentist_Data_LIMIT <- Dentist_Data |>
  filter(Sum.salgsi < 100000000, DistCBD < 250,
         Driftsres < 20000000, Driftsres > -10000000, DistMal < 80)
```

```
Dentist_Data_LIMIT |>
  ggplot(aes(x = DistCBD, y = Sum.salgsi, colour = DistCBD)) +
  geom_point(lwd = .9) + labs(x = "Distance from CBD", y = "Income") +
  theme_bw()
```

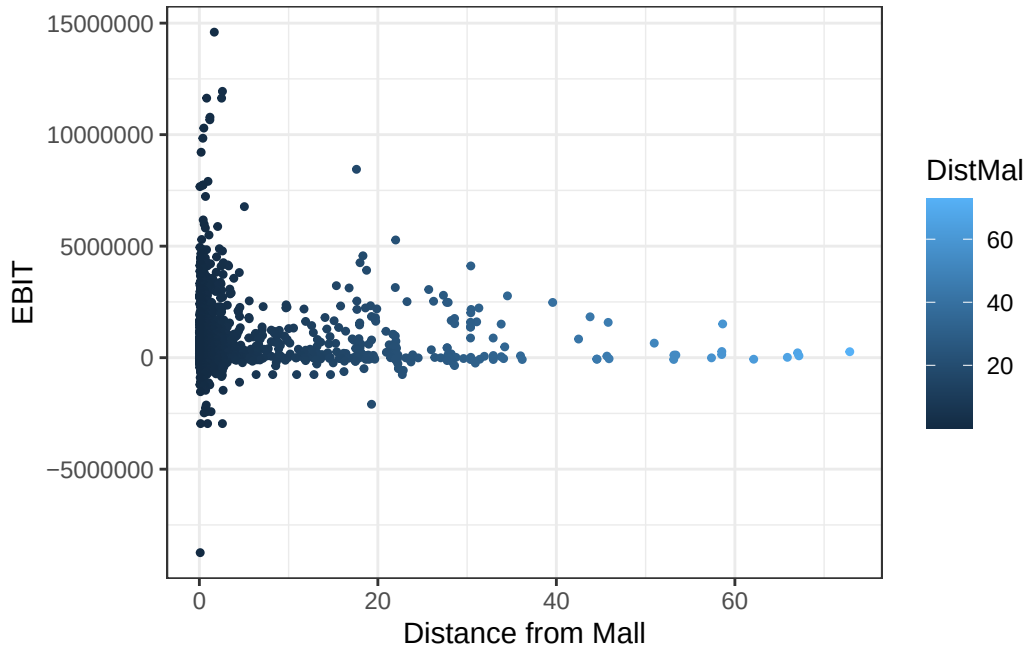



```
Dentist_Data |>
  ggplot(aes(x = DistMal, y = Driftsres, colour = DistMal)) +
  geom_point(lwd = .9) +
  labs(x = "Distance from Mall", y = "EBIT") +
  theme_bw()
```

```
Warning: Removed 3838 rows containing missing values (geom_point).
```




```
Dentist_Data_LIMIT |>
  ggplot(aes(x = DistMal, y = Driftsres, colour = DistMal)) +
  geom_point(lwd = .9) +
  labs(x = "Distance from Mall", y = "EBIT") +
  theme_bw()
```




```
lm1 <- lm(Sum.salgsi ~ X_sum, data = Dentist_Data_LIMIT)
lm2 <- lm(Sum.salgsi ~ DistCBD + DistMal + X_sum, data = Dentist_Data_LIMIT)
```

```
huxreg(list("Model 1" = lm1, "Model 2" = lm2), statistics = c(N = "nobs", R2 = "r.squared"), const. =
```

Econometric approach

	Model 1	Model 2
(Intercept)	3951211.459 ***	4231549.685 ***
	(189925.810)	(239690.185)
X_sum	14.742 *	9.705
	(6.258)	(6.779)
DistCBD		-4773.850
		(4538.911)
DistMal		-19483.503
		(17864.173)
N	1801	1801
R2	0.003	0.005

Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$ T statistics in brackets.