MSB106 - Assignment

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

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

```r
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

```r
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)
Dist_Comp_Den <- read.csv("Dist_Comp_Den.csv") |>
  select(fid, distance)
NOR_KOMM <- read_csv("NOR_KOMM.csv")
NOR_KOMM_2 <- read_csv("NOR_KOMM_2.csv")
```

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

```r
Commune_Data <- inner_join(Commune_Data, NOR_OSM_SHOP_MAL, by = "fid") |>
  select(knr, DistMal, fid, Turnover_capita_retail, shop, Vinmonopolet) |>
  rename("SizeMall" = shop, "Winemonopoly" = Vinmonopolet)
```

```r
Commune_Data$SizeMall <- as.numeric(Commune_Data$SizeMall)
```

```
Warning: NAs introduced by coercion
```

```
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, Hub
  rename("DistMal" = HubDist.x, "DistCBD" = HubDist.y, "HubNameMal" = HubName.x, "HubNameCBD" = HubNa
```

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

```
Dentist_Data <- inner_join(Dentist_Data, Dist_Comp_Den, by = "fid") |>
  rename("DistComp" = distance)
```

```
Dentist_Data <- left_join(Dentist_Data, NOR_OSM_SHOP_MAL, by = "fid") |>
  select(fid, Juridisk.n, Antall.ans, Sum.Drifts, Sum.salgsi, latlong, DistCBD, DistMal, X_sum, DistC
  rename("SizeMall" = shop, "Winemonopoly" = Vinmonopolet)
```

```
Dentist_Data$SizeMall <- as.numeric(Dentist_Data$SizeMall)
```

```
Warning: NAs introduced by coercion
```

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	DistMal	fid	Turnover_capita_retail	SizeMall	Winemonopoly
Length:435	Min. : 0.01854	Min. : 6.0	Min. : 17366	Min. : 2.00	Length:435
Class :character	1st Qu.: 9.41403	1st Qu.:220.0	1st Qu.: 59250	1st Qu.: 13.00	Class :character
Mode :character	Median :18.87373	Median :291.0	Median : 83203	Median : 25.00	Mode :character
NA	Mean :23.82982	Mean :280.8	Mean : 84719	Mean : 37.22	NA
NA	3rd Qu.:32.67127	3rd Qu.:363.0	3rd Qu.:106257	3rd Qu.: 50.00	NA
NA	Max. :98.37355	Max. :424.0	Max. :218728	Max. :206.00	NA
NA	NA	NA	NA	NA's :124	NA

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

| Antall.ans	Sum.Drifts	Sum.salgsi	latlong
Min. : 0.00	Min. :0.000e+00	Min. : -28000	Length:5740
1st Qu.: 1.00	1st Qu.:1.117e+06	1st Qu.: 658250	Class :character
Median : 1.00	Median :3.616e+06	Median : 3176500	Mode :character
Mean : 39.26	Mean :5.085e+07	Mean : 50123495	NA
3rd Qu.: 3.00	3rd Qu.:6.820e+06	3rd Qu.: 6248250	NA
Max. :6870.00	Max. :1.058e+09	Max. :1058102000	NA
NA	NA's :3838	NA's :3838	NA

```r
kable(summary(Dentist_Data[7:10]))
```

DistCBD	DistMal	X_sum	DistComp
Min. : 0.0046	Min. : 0.0008	Min. : 0	Min. : 0.0
1st Qu.: 1.3912	1st Qu.: 0.3145	1st Qu.: 6129	1st Qu.: 0.0
Median : 6.7067	Median : 0.9164	Median :14602	Median : 0.0
Mean : 20.5000	Mean : 4.6331	Mean :22455	Mean : 614.4
3rd Qu.: 19.8885	3rd Qu.: 2.5220	3rd Qu.:28904	3rd Qu.: 197.7
Max. :470.1005	Max. :105.8867	Max. :92591	Max. :60577.8

```r
kable(summary(Dentist_Data[11:12]))
```

| SizeMall	Winemonopoly
Min. : 1.00	Length:5740
1st Qu.: 13.00	Class :character
Median : 29.00	Mode :character
Mean : 40.53	NA
3rd Qu.: 55.00	NA
Max. :206.00	NA
NA's :5515	NA

```r
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()
```
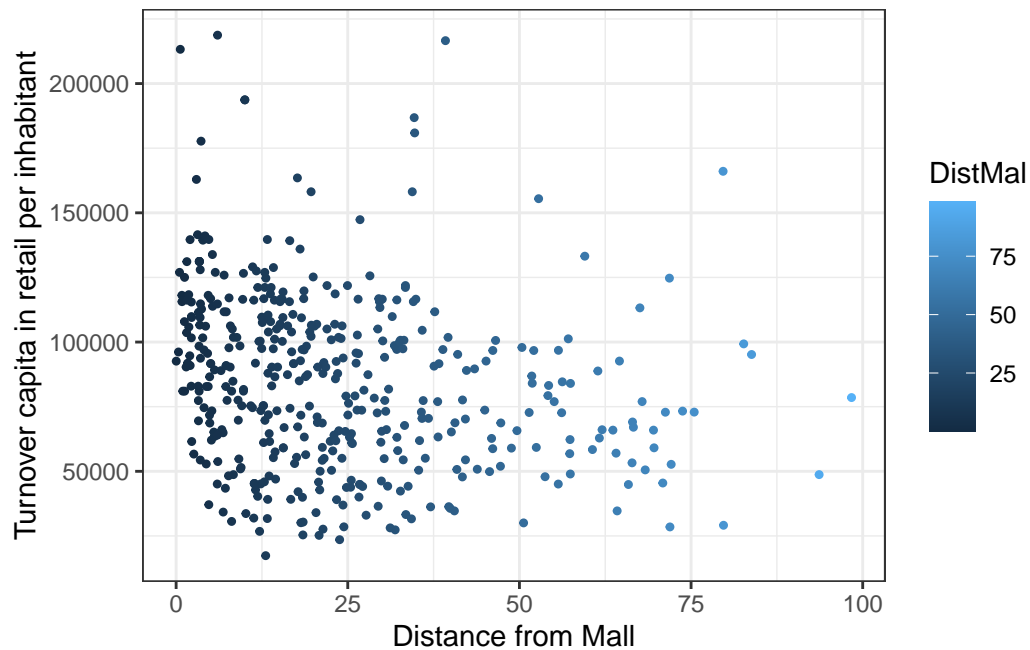
Might be more appropriate to explore the distance from CBDs to Malls compared to turnover capita in retail per inhabitant, rather than the distance from the geographical centroid of Norwegian Communes, because of the tendency of bigger population density around CBDs.
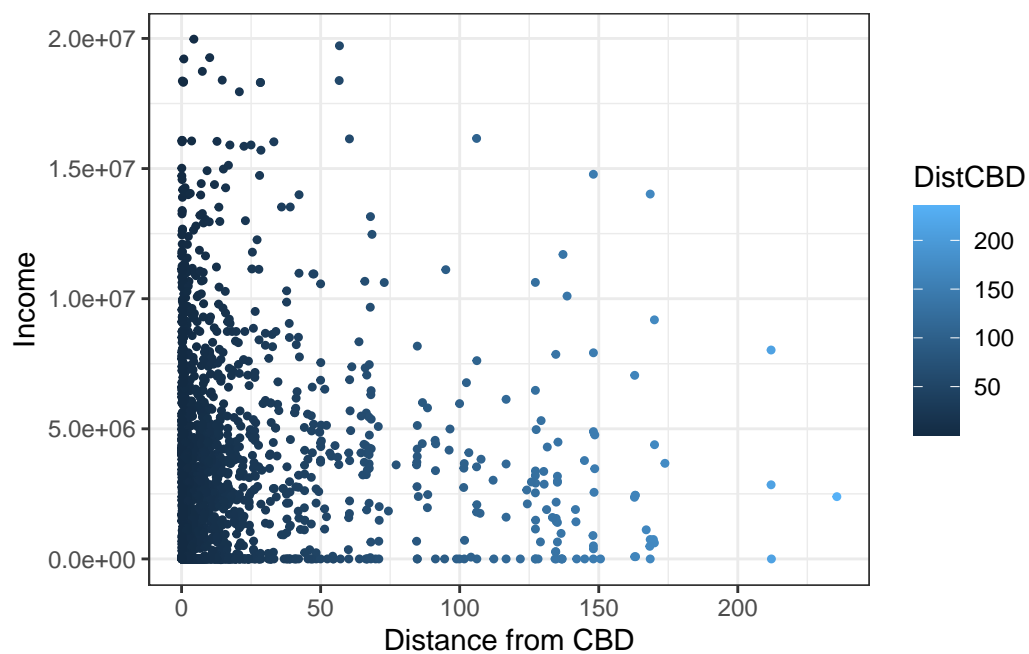
```
Dentist_Data |>
  ggplot(aes(x = DistCBD, y = Sum.salgsi, colour = DistCBD)) +
  geom_point(lwd = .9) +
  labs(x = "Distance from CBD", y = "Income") +
  theme_bw()
```

Warning: Removed 3838 rows containing missing values (geom_point).
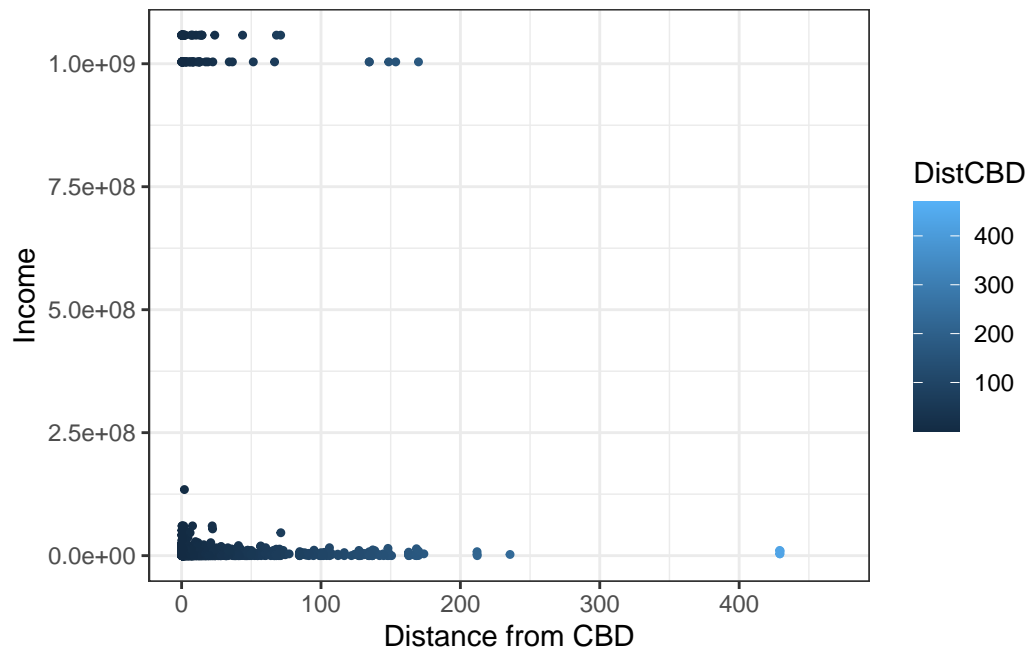
```
Dentist_Data_LIMIT <- Dentist_Data |>
  filter(Sum.salgsi < 100000000, DistCBD < 250,
         Sum.Drifts < 20000000, Sum.Drifts > -10000000, DistMal < 80)
```
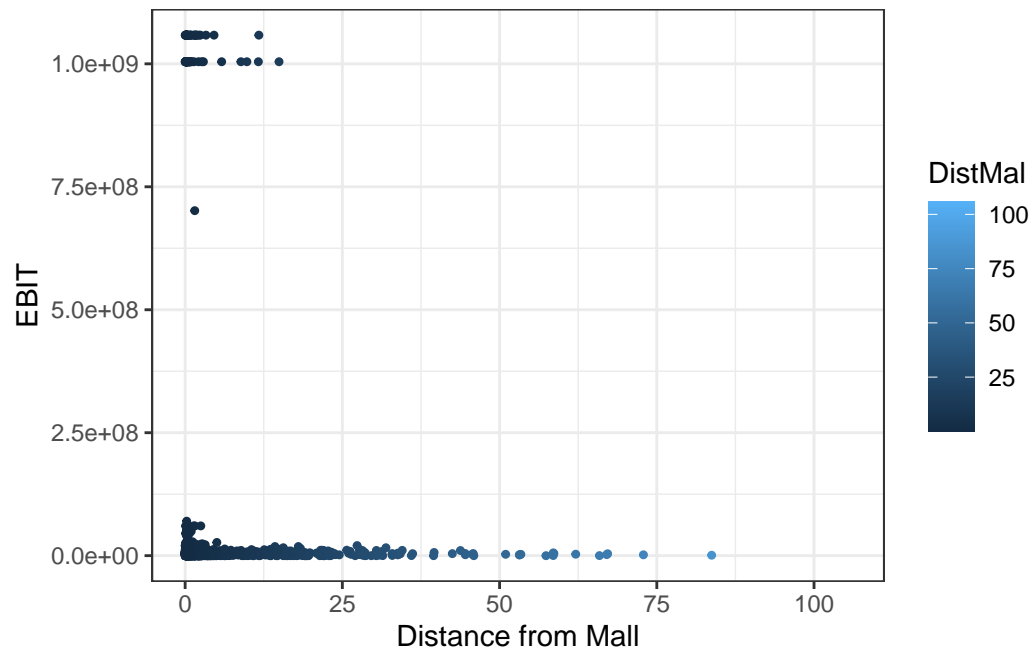
```
Dentist_Data_LIMIT |>
  ggplot(aes(x = DistCBD, y = Sum.salgsi, colour = DistCBD)) +
  geom_point(lwd = .9) + labs(x = "Distance from CBD", y = "Income") +
  theme_bw()
```
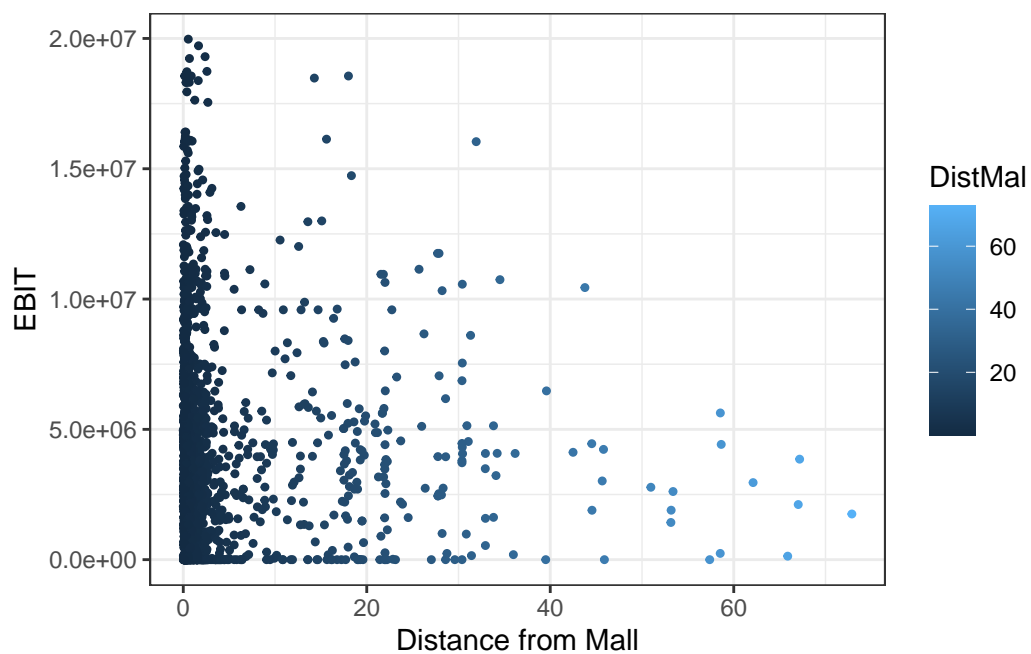
```
Dentist_Data |>
  ggplot(aes(x = DistMal, y = Sum.Drifts, colour = DistMal)) +
  geom_point(lwd = .9) +
  labs(x = "Distance from Mall", y = "EBIT") +
  theme_bw()
```
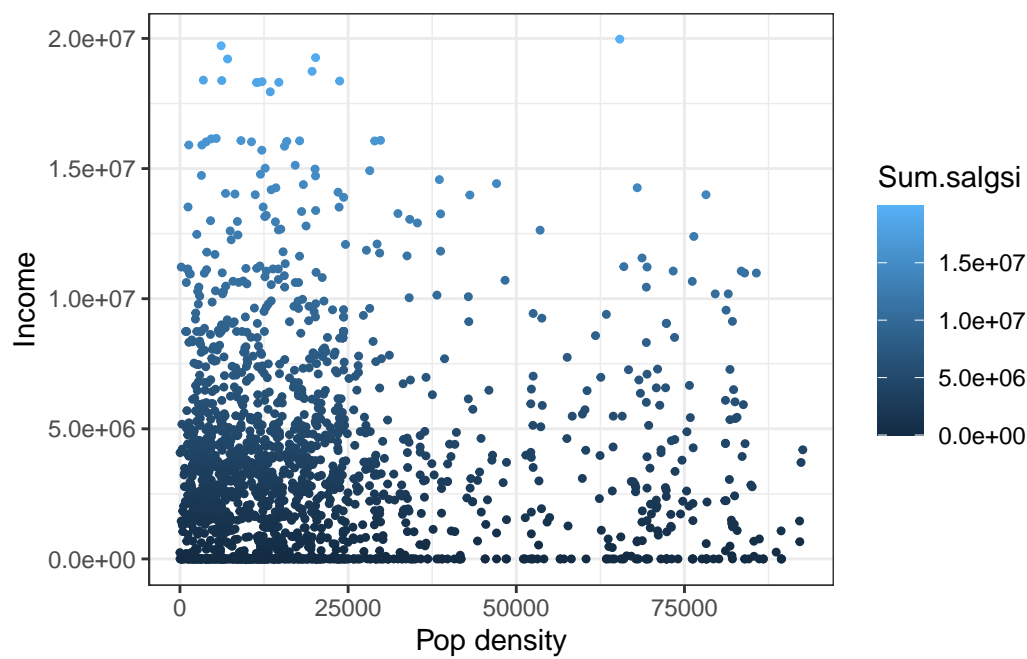
Warning: Removed 3838 rows containing missing values (geom_point).

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

```
Dentist_Data_LIMIT |>
  ggplot(aes(x = X_sum, y = Sum.salgsi, colour = Sum.salgsi)) +
  geom_point(lwd = .9) +
  labs(x = "Pop density", y = "Income") +
  theme_bw()
```

```
lm1 <- lm(Sum.salgsi ~ DistComp + DistMal + SizeMall + (DistMal*SizeMall) + DistCBD + X_sum + (DistCI
```

```
huxreg(list("Dentist Sales" = lm1), statistics = c(N = "nobs", R2 = "r.squared"), const. = "(Intercep
```

Econometric approach

	Dentist Sales
(Intercept)	6762625.038 ***
	(723928.279)
DistComp	-622.876
	(904.073)
DistMal	6654.475
	(51523.559)
SizeMall	10860.864
	(9589.931)
DistCBD	-3499.524
	(19661.750)
X_sum	19.729
	(18.522)
DistMal:SizeMall	-753.373
	(711.846)
DistCBD:X_sum	1.195
	(2.148)
N	187
R2	0.030

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