**ENDRER DATETIME OGSÅ SETTER DEN TIL DEN 15 HVER MÅNED.**

df['måned'] = pd.to\_datetime(df['måned'], format='%YM%m')

df['måned'] = df['måned'] + pd.to\_timedelta(14, unit='d')

df

df["ulykkestype"] = df["ulykkestype"].str.replace('\xa0', "") **Fjerner xa0**



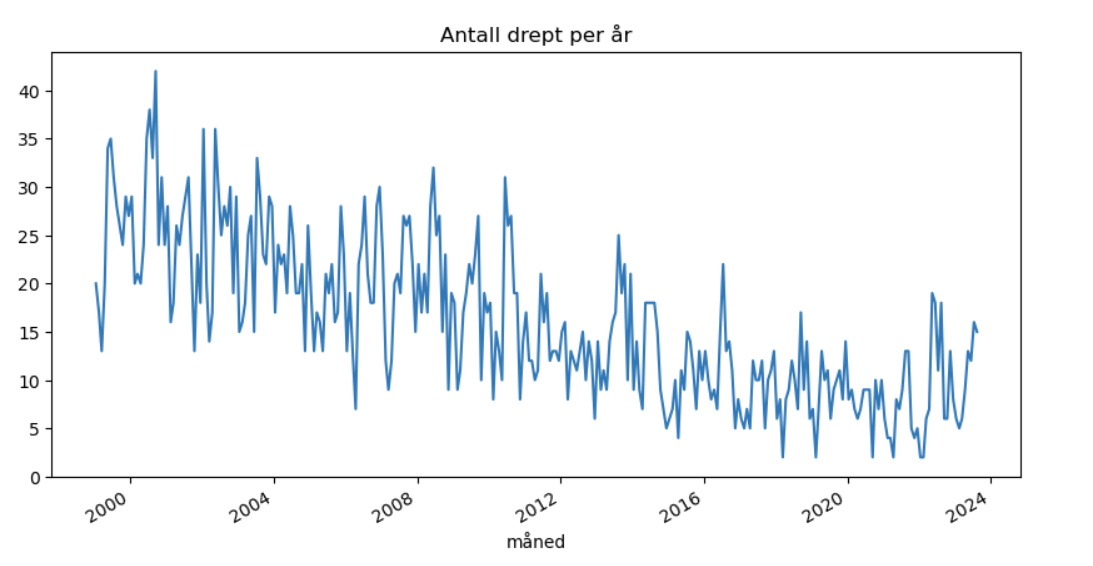
df["baremåned"] = df["måned"].dt.year

df

drept\_data = df[df["skadegrad"] == "Drept"].groupby("måned")["08329: Drepte eller skadde i trafikkulykker,"].sum()

drept\_data

drept\_data.plot(x="år", y="08329: Drepte eller skadde i trafikkulykker,", kind="line", figsize=(10, 5), title="Antall drept per måned")



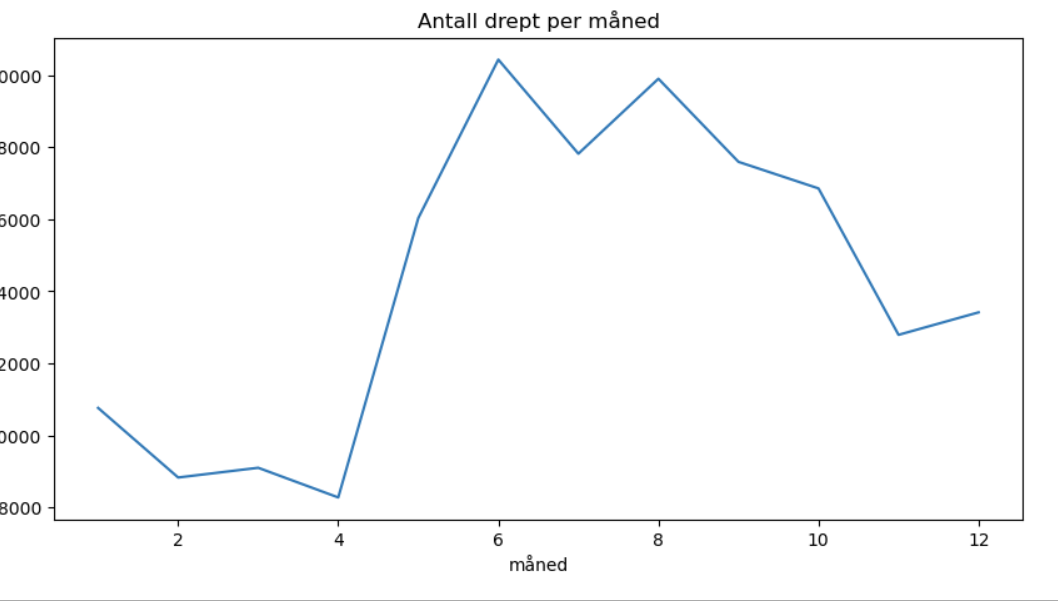
**Viser endringen over år av antall drepte.**

**C)**

**ikke\_drept\_data = df[df["skadegrad"] != "Drept"]**

**månedlig\_ikke\_drept = ikke\_drept\_data.groupby(df["måned"].dt.month)["08329: Drepte eller skadde i trafikkulykker,"].sum()**

månedlig\_ikke\_drept.plot(x="måned", y="Antall drept", kind="line", figsize=(10, 5), title="Antall drept per måned")

****

**D)**

**df["baremåned"] = df["måned"].dt.year**

**drept\_data = df[df["skadegrad"] != "Skadde i alt"]**

**pivot\_table = drept\_data.pivot\_table(index="måned", columns="skadegrad", values="08329: Drepte eller skadde i trafikkulykker,", aggfunc="sum")**

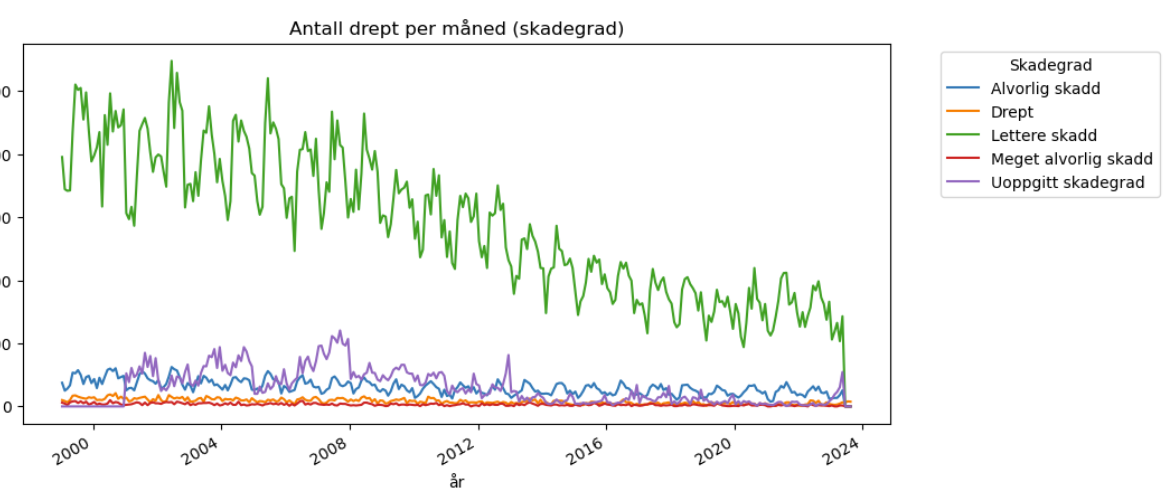
**pivot\_table.plot(kind="line", figsize=(10, 5))**

**plt.title("Antall drept per måned (skadegrad)")**

**plt.xlabel("år")**

**plt.ylabel("Antall drept eller skadde")**

**plt.legend(title="Skadegrad", bbox\_to\_anchor=(1.05, 1), loc='upper left')**

****

**E)**

**df["baremåned"] = df["måned"].dt.year**

**drept\_data = df[(df["ulykkestype"] == "C. Møting ved forbikjøring") | (df["ulykkestype"] == "D. Andre møteulykker")]**

**pivot\_table = drept\_data.pivot\_table(index="måned", columns="ulykkestype", values="08329: Drepte eller skadde i trafikkulykker,", aggfunc="sum")**

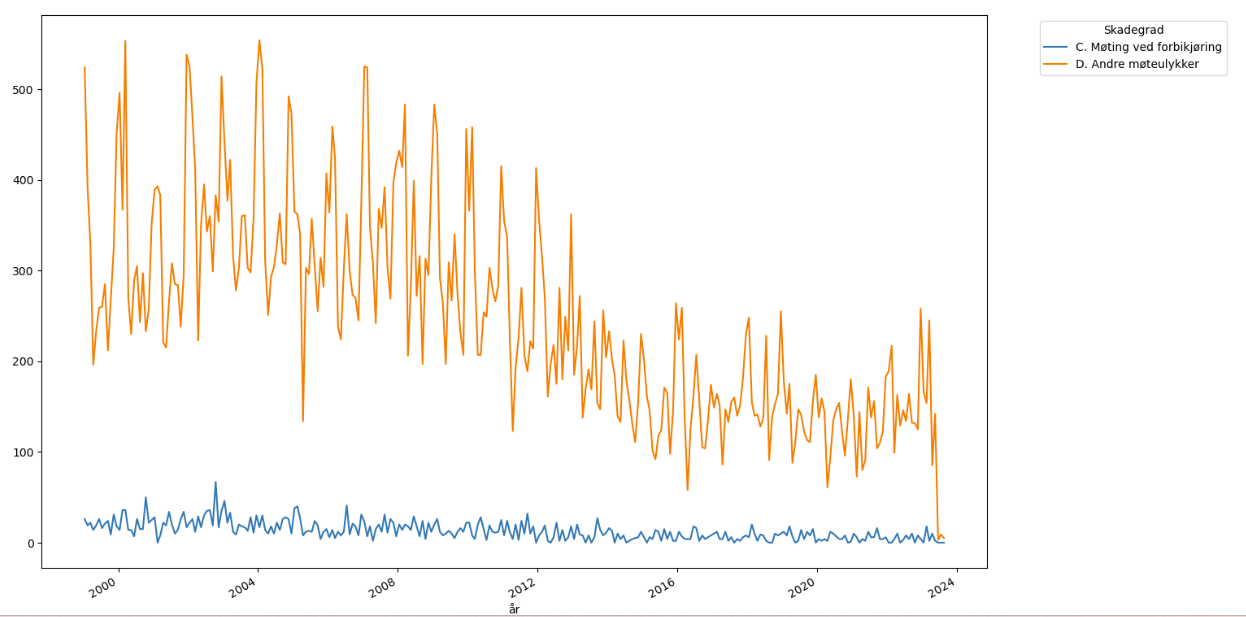
**pivot\_table.plot(kind="line", figsize=(15, 10))**

**plt.title("")**

**plt.xlabel("år")**

**plt.ylabel("")**

**plt.legend(title="Skadegrad", bbox\_to\_anchor=(1.05, 1), loc='upper left')**

****

**Oppgave 2:**

**A)**

**library(lubridate)**

**parquet\_file <- "C:/onemin2015/solar.parquet"**

**df <- arrow::read\_parquet(parquet\_file)**

**df$TIMESTAMP <- lubridate::with\_tz(df$TIMESTAMP, tzone = "America/New\_York")**

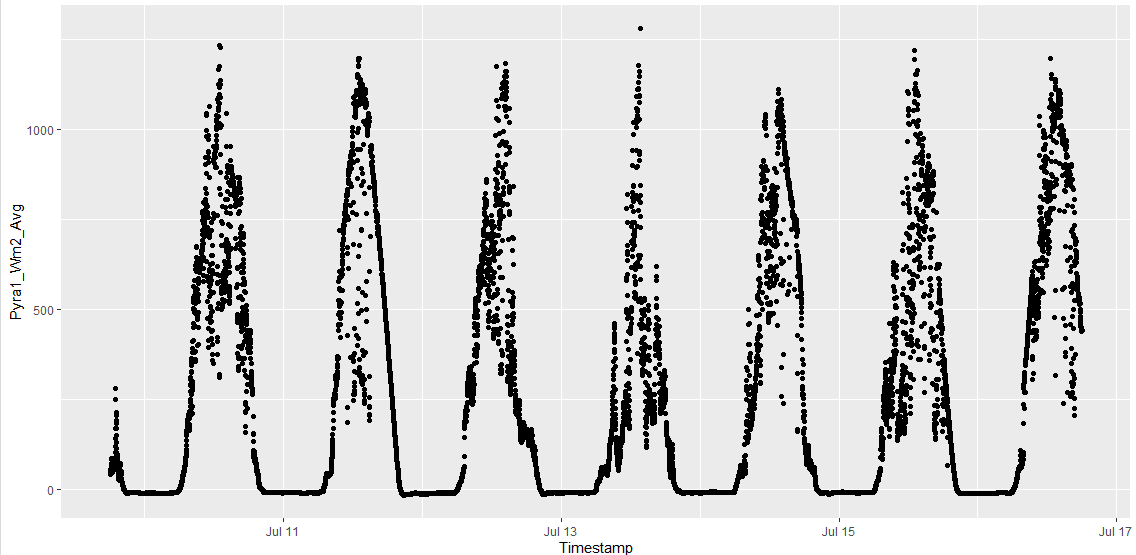
**library(ggplot2)**

**df$TIMESTAMP <- as.POSIXct(df$TIMESTAMP)**

**start\_date <- as.POSIXct("2015-06-01") end\_date <- as.POSIXct("2015-08-31")**

**df\_filtered <- df[df$TIMESTAMP >= start\_date & df$TIMESTAMP <= end\_date, ]**

**ggplot(df\_filtered, aes(x = TIMESTAMP, y = Pyra1\_Wm2\_Avg)) + geom\_point() + labs(x = "Timestamp", y = "Pyra1\_Wm2\_Avg") + ggtitle("Plot of Pyra1\_Wm2\_Avg Over Time (June 2015 - August 2015)")**

****

B)

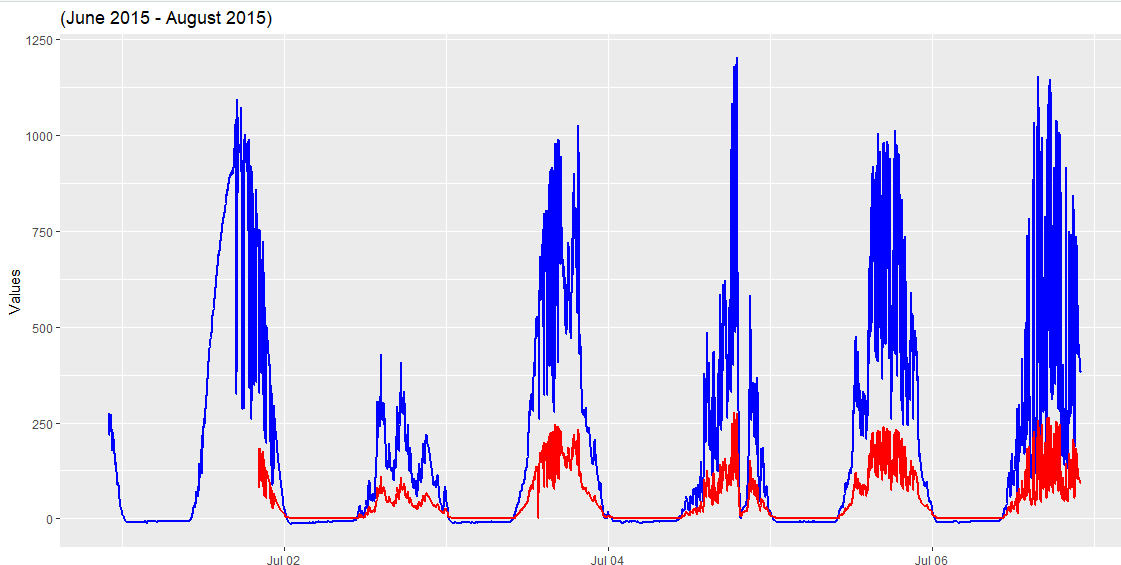
> df$TIMESTAMP <- as.POSIXct(df$TIMESTAMP)

> start\_date <- as.POSIXct("2015-07-01")

> end\_date <- as.POSIXct("2015-07-7")

> df\_filtered <- df[df$TIMESTAMP >= start\_date & df$TIMESTAMP <= end\_date, ]

> ggplot(df\_filtered, aes(x = TIMESTAMP)) + geom\_line(aes(y = Pyra1\_Wm2\_Avg), color = "blue", size = 1) + geom\_line(aes(y = InvPDC\_kW\_Avg), color = "red", size = 1) + labs(x = "Timestamp", y = "Values") + ggtitle("(June 2015)")



C)   
  
#når jeg brukte correlation <- cor(df$Pyra1\_Wm2\_Avg, df$InvPDC\_kW\_Avg) fikk jeg Pearson Correlation between Pyra1\_Wm2\_Avg and InvPDC\_kW\_Avg: NA

missing\_values <- sum(is.na(df$Pyra1\_Wm2\_Avg) | is.na(df$InvPDC\_kW\_Avg))

df\_cleaned <- df[complete.cases(df$Pyra1\_Wm2\_Avg, df$InvPDC\_kW\_Avg), ]

correlation <- cor(df\_cleaned$Pyra1\_Wm2\_Avg, df\_cleaned$InvPDC\_kW\_Avg)

cat("Pearson Correlation between Pyra1\_Wm2\_Avg and InvPDC\_kW\_Avg:", correlation, "\n")



#Ettersom den lander på rundt 0.03 vil det si at de er lite avhengige av hverandre og det er svak korrelasjon, det er noe korrelasjon, men den er ikke sterk. (dette er tatt fra hele året ikke en uke)

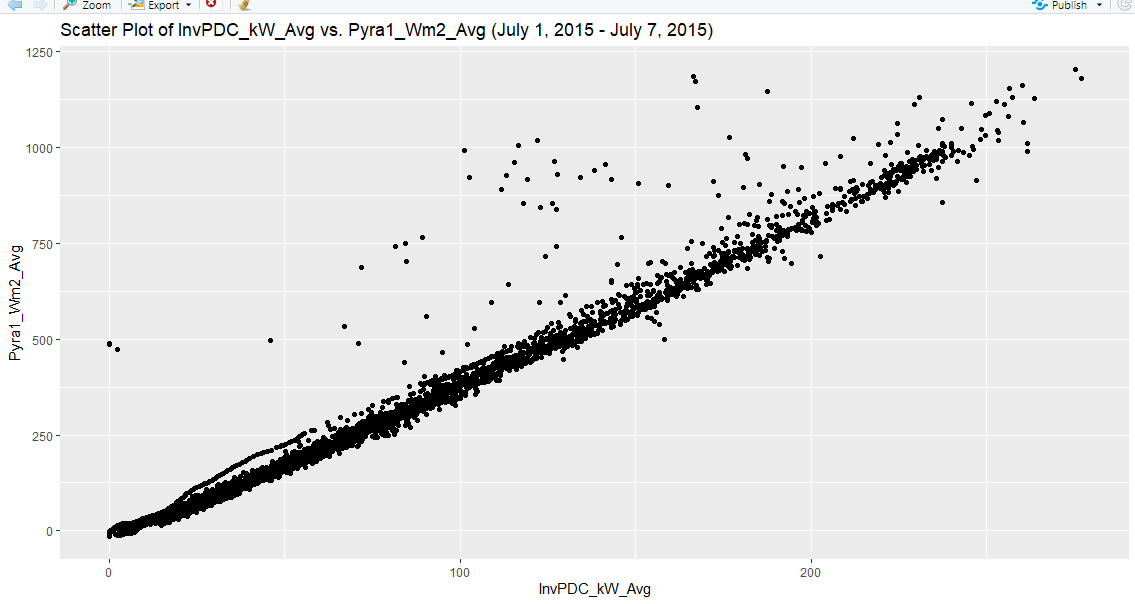
D)

start\_date <- as.POSIXct("2015-07-01")

end\_date <- as.POSIXct("2015-07-07")

df\_filtered <- df[df$TIMESTAMP >= start\_date & df$TIMESTAMP <= end\_date, ]

ggplot(df\_filtered, aes(x = InvPDC\_kW\_Avg, y = Pyra1\_Wm2\_Avg)) + geom\_point() + labs(x = "InvPDC\_kW\_Avg", y = "Pyra1\_Wm2\_Avg") + ggtitle("July 1, 2015 - July 7, 2015)")



E) start\_date <- as.POSIXct("2015-01-01")

end\_date <- as.POSIXct("2015-12-31")

df\_filtered\_2015 <- df[df$TIMESTAMP >= start\_date & df$TIMESTAMP <= end\_date, ]

model <- lm(InvPDC\_kW\_Avg ~ Pyra1\_Wm2\_Avg, data = df\_filtered\_2015)

summary(model)

#Ettersom vi brukte hele året og ikke bare en uke, viser det at modellen ikke er en god modell. i motsetening om vi tar kun en uke som i de forrige oppgavene virker det som en veldig god modell. Dette ser vi ved bruk av Multiple R squared som ligger på 0.0009301. Men om vi tar tidsperioden 1 juli til 7 juli får vi Multiple R-squared: 0.9787 noe som vi sier at modellen er bra på ukes basis, ikke årsbasis.

#> conf\_interval <- confint(model, level = 0.95)

> conf\_interval

#(Intercept) -143.31515158 -140.61408878

Pyra1\_Wm2\_Avg 0.04028549 0.04877463