Assignment 3

#Assignment 3

Task 1

The following code is used to extract the difference between the left join of the company data (AS_merged) and the exploration data (expl_oper)

```
library(arsenal)
```

```
##
## Attaching package: 'arsenal'
```

```
## The following object is masked from 'package:lubridate':
##
##     is.Date
```

```
expl_AS_left_join <- expl_oper %>% left_join(AS_merged, by=c("NPD_id", "year"))

diff_set <- setdiff(expl_AS_left_join, expl_AS)

nrow(diff_set)
```

```
## [1] 93
```

Why were they dropped?

The diference is observed by taking the difference of a inner join of the variables NPD_id and year, and a left join of the same variables. 93 observations thus appear in the exploratory dataset that does not occur in the company dataset For a left join / merge between two rows to be made both the NPD_id and year much find an equal pairing in the other data set.

The loss of data appears random and this may limit the extent of the problems caused by the lacking of data in the company dataset

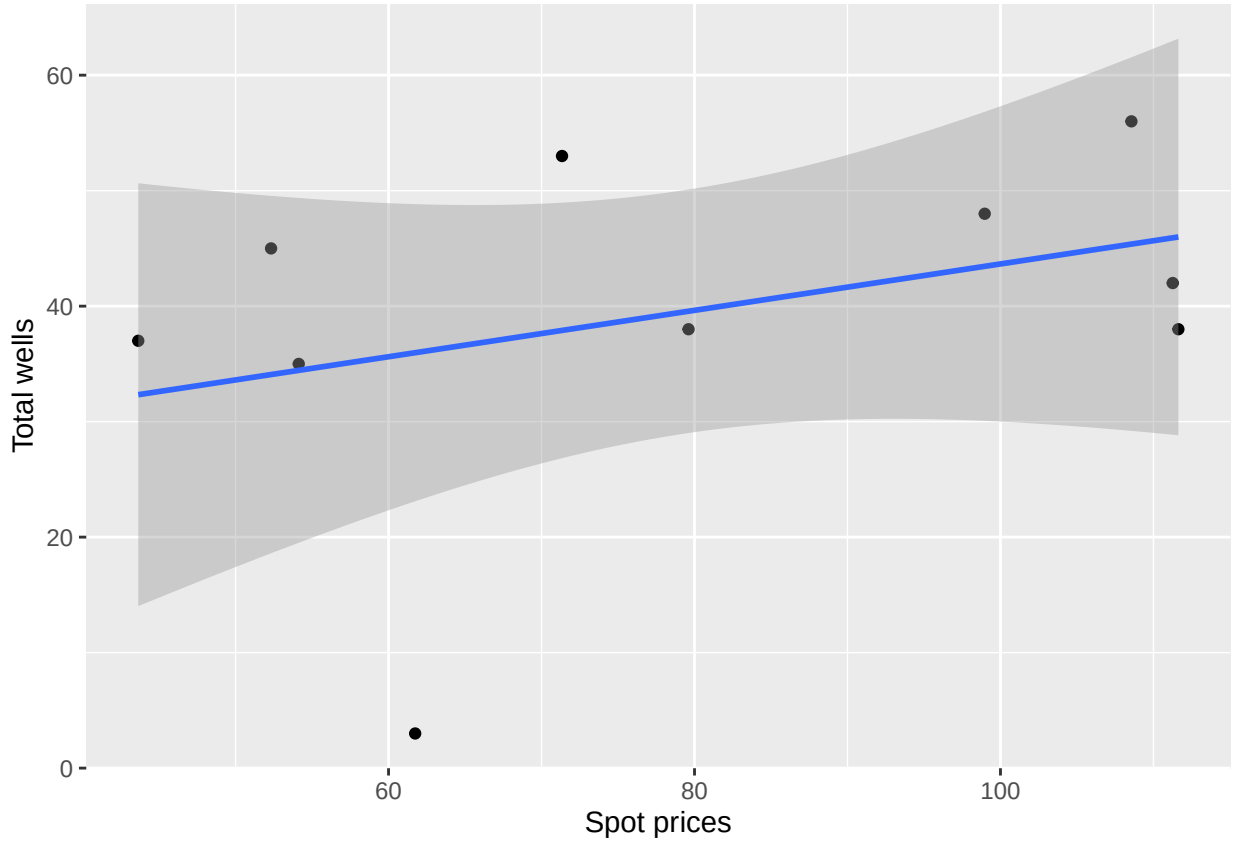
What do they have in common?

TASK 2

We run a linear regression of

```r
brent_spotprices <- read.csv("Europe_Brent_Spot_Price_FOB.csv") %>% tail(., -4)

brent_spotprices %<>%
  mutate(year = as.numeric(rownames(brent_spotprices)),
         Europe.Brent.Spot.Price.FOB = as.numeric(Europe.Brent.Spot.Price.FOB)) %>%
  rename(spot_price = Europe.Brent.Spot.Price.FOB)


explAGG_spot <- explAgg_AS %>% inner_join(brent_spotprices, by = c("year"))

explAGG_spot %>% group_by(spot_price) %>%
  summarise(total_n_wells = sum(numbWells)) %>%
  ggplot(aes(x= spot_price, y= total_n_wells)) +
  geom_point() +
  geom_smooth(method = "lm") +
  xlab("Spot prices") +
  ylab("Total wells")
```

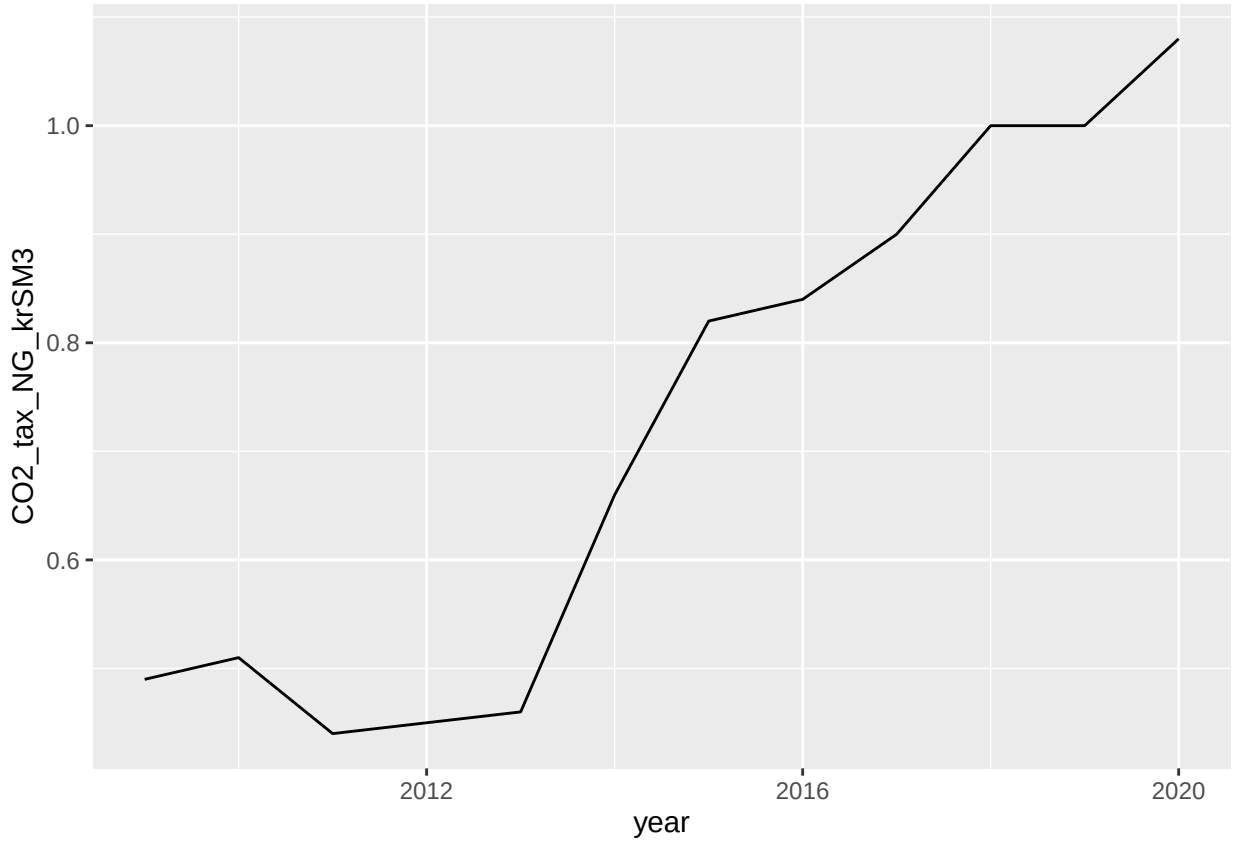
```
## `geom_smooth()` using formula 'y ~ x'
```



TASK 3

Overview of the CO2 taxes in the period 2009-2020.

```
CO2_tax = tibble(
  year=2009:2020,
  CO2_tax_NG_krSM3=c(0.49, 0.51, 0.44, 0.45, 0.46, 0.66,.82, 0.84, 0.90, 1, 1, 1.08)
  )

CO2_tax %>% ggplot(aes(x=year, y=CO2_tax_NG_krSM3)) +
  geom_line()
```

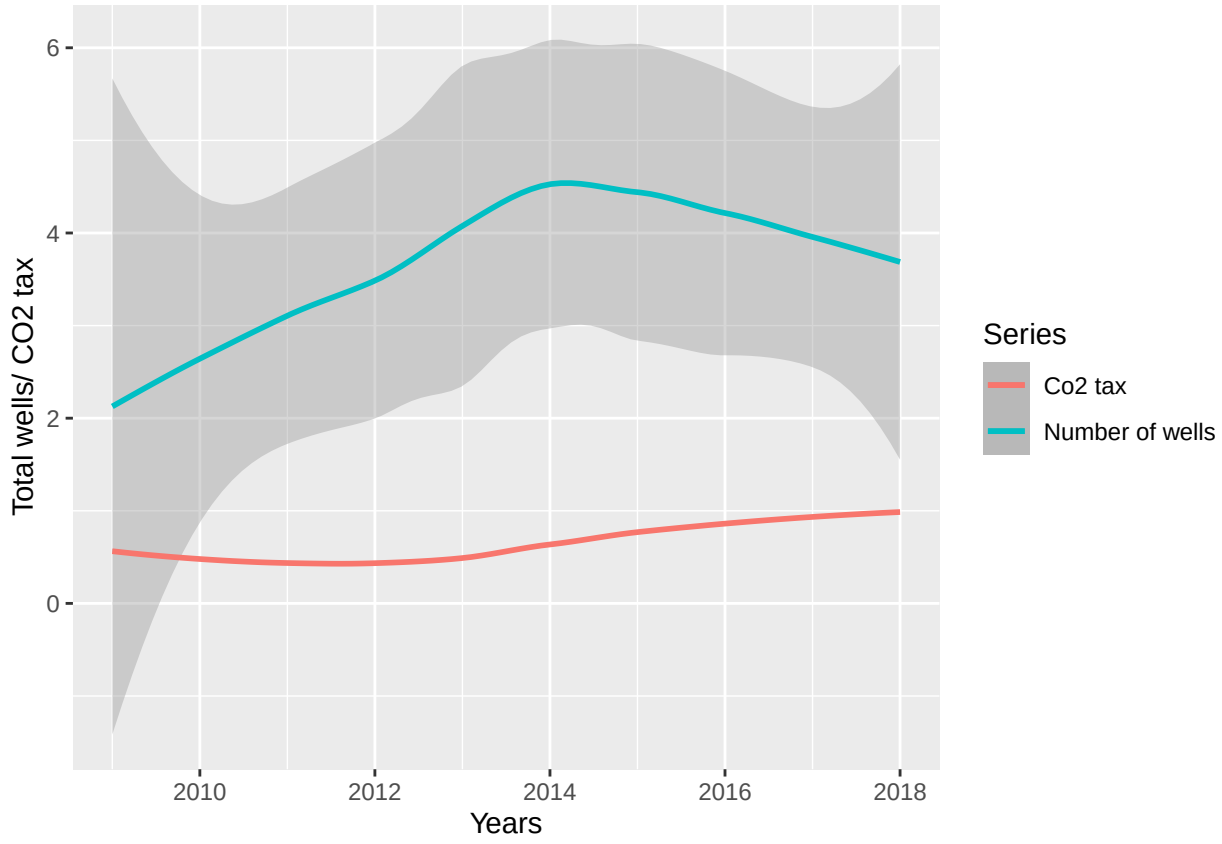


The exploratory drilling seems to be unaffected by the hike in the Co2 taxes. Simply plotting the number of exploratory wells against the the increase Co2 tax levels, does not seem to correspond with the rate or time period in which the number of exploratory wells increase

```
## Run regression tests

num_wells <- left_join(explAGG_spot, CO2_tax, by = "year")
num_reg <- lm(numbWells ~ CO2_tax_NG_krSM3, data = num_wells)




## Plot CO2 tax vs exploration

num_wells %>%
  ggplot(aes(x = year, group = 1)) +
  geom_smooth(aes(y = numbWells, color = "Number of wells")) +
  geom_smooth(aes(y = CO2_tax_NG_krSM3, color = "Co2 tax")) +
```

```
  xlab("Years") +
  ylab("Total wells/ CO2 tax") +
  labs(colour = "Series")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Conducting a simple linear regression tests suggest that the is little correlation betweeen

```
summary(num_reg)
```

```
##
## Call:
## lm(formula = numbWells ~ CO2_tax_NG_krSM3, data = num_wells)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1415 -2.4305 -1.6938  0.4444 19.5695
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)         2.825      1.367   2.067   0.0412 *
## CO2_tax_NG_krSM3    1.317      1.976   0.666   0.5067
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.328 on 105 degrees of freedom
## Multiple R-squared:  0.00421,    Adjusted R-squared:  -0.005274
## F-statistic: 0.4439 on 1 and 105 DF,  p-value: 0.5067
```

Task 4

The following regression is conducted

```
reg1 = lm(numbWells ~ total_assets + profitability, data=explAgg_AS)
summary(reg1)
```

```
##
## Call:
## lm(formula = numbWells ~ total_assets + profitability, data = explAgg_AS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.8817 -1.9410 -1.3936  0.4946 11.2441
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.428e+00  3.798e-01   6.391 5.13e-09 ***
## total_assets   2.673e-08  3.491e-09   7.656 1.18e-11 ***
## profitability -4.859e-03  1.015e-02  -0.479    0.633
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.47 on 101 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.3789, Adjusted R-squared:  0.3667
## F-statistic: 30.81 on 2 and 101 DF,  p-value: 3.572e-11
```

Total_assets is more significant than profitability. The R-squared is quite low.