DA3: selecting columns

We are practicing selecting columns from our data and then creating new columns and then using them for plotting.

In [ ]:
```
import pandas as pd
from plotnine import *

infections = pd.read_csv("data/infections.csv")

infections.columns
```

Scale the 'crp_level' column manually creating a new column

We use the **.mean()** and **.std()** methods to calculate mean & standard deviation for scaling.

In [ ]:
```
infections['crp_scaled'] = (infections['crp_level'] - infections['crp_level'].mean()) / infections['crp_level'].std()
```

NB: In geom_jitter the width parameter controls the horizontal jitter. And in geom_hline we use yintercept to position the horizontal and the color and size parameters to set the style of the line.

In [ ]:
```
p = (
    ggplot(infections, aes(x='quarter', y='crp_scaled')) +
    geom_jitter(width=0.1) +
    geom_hline(yintercept=0, color="blue", size=1)
)

p.show()
```

mutate(infections, risk_factor = ifelse(vaccination_status == "unvaccinated" & icu_admission == TRUE, "high", NA))

In [ ]:
```
conditions = [
    (infections["vaccination_status"] == "unvaccinated") & (infections["icu_admission"] == True) ]

choices =  ['high']
infections["risk_factor"] = np.select(conditions, choices, default="")
```

In [ ]:
```
# to display in a nice table all entries including new column
from IPython.display import display

# Display DataFrame with all columns
display(infections)
```

In [ ]:
```

```