

Performance Paradox

load the data and packages

```
library(tidyverse)
```

```
## Warning: package 'lubridate' was built under R version 4.3.1
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.2      v readr      2.1.4
```

```
## v forcats    1.0.0      v stringr    1.5.0
```

```
## v ggplot2    3.4.2      v tibble     3.2.1
```

```
## v lubridate  1.9.3      v tidyr      1.3.0
```

```
## v purrr      1.0.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
df <- read.csv("/Users/kaz/Desktop/MMA - WINTER Code/Talent Analytics/Session3/performance_data.csv")
```

```
head(df)
```

```
##   day   worker1 w1_intervention   worker2 w2_intervention   worker3
## 1    1 -1.2849391             None 0.638546576             None 0.642798502
## 2    2  1.4106987             None -0.066419765             None 0.366053529
## 3    3  0.3531135             None 0.024467507             None 0.553280729
## 4    4  0.3604826             None -0.835858608             None 1.305998180
## 5    5 -0.6501855             None -0.004233503             None -0.005802895
## 6    6  1.7144270             None -1.158510709             None -0.441555347
##   w3_intervention
## 1                None
## 2                None
## 3                None
## 4                None
## 5                None
## 6                None
```

average performance of each worker

```
df %>%
  summarise(worker1 = mean(worker1),
            worker2 = mean(worker2),
            worker3 = mean(worker3))
```

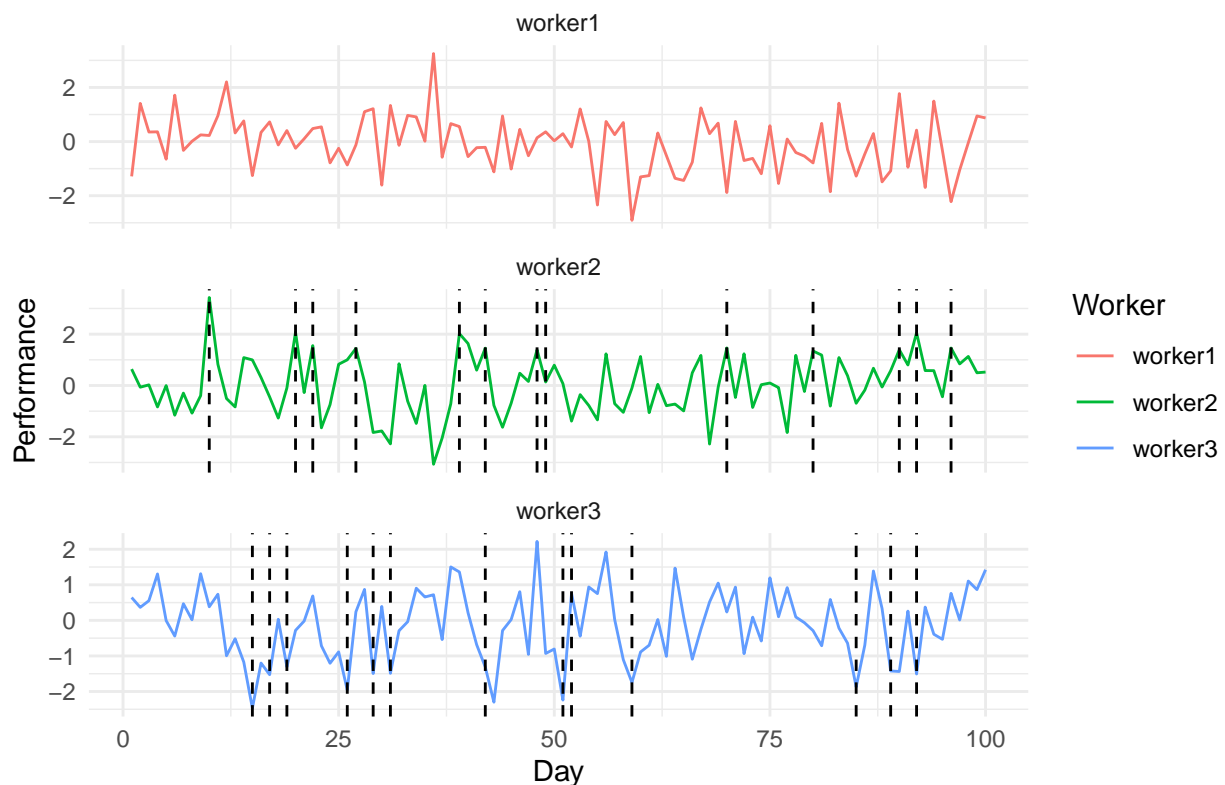
```
##      worker1      worker2      worker3
## 1 -0.0621415  0.03260368 -0.1198092
```

visualize the data

```
long_df <- df %>%
  gather(key = "worker", value = "performance", worker1, worker2, worker3) %>%
  mutate(intervention = ifelse(worker == "worker1" & w1_intervention != "None", as.character(day),
                              ifelse(worker == "worker2" & w2_intervention != "None", as.character(day),
                                      ifelse(worker == "worker3" & w3_intervention != "None", as.character(day),
                                              "None"))))

# Create the line plot with separate facets for each worker, aligned vertically
ggplot(long_df, aes(x = day, y = performance, group = worker, colour = worker)) +
  geom_line() +
  geom_vline(data = subset(long_df, !is.na(intervention)), aes(xintercept = as.numeric(intervention)),
            linetype = "dashed", colour = "black", size = 1) +
  facet_wrap(~ worker, scales = "free_y", nrow = 3) +
  labs(title = "Performance over Days with Interventions",
       x = "Day",
       y = "Performance",
       colour = "Worker") +
  theme_minimal()
```

Performance over Days with Interventions



Notes

- One intervention happens when performances are low and the other happens when performances are high
- rewards vs punishment

SO which is better?

- on average, worker2 has a higher performance than worker3
- worker2 is motivated or “rewarded” when they perform well
- worker3 is motivated or “punished” when they perform poorly

In the worker2 graph, we can see that their performance drops after the intervention. In the worker3 graph, we can see that their performance increases after the intervention.

In conclusion, I am not sure which intervention is better. It depends on types of work and behaviour of workers