Exercise 1 Group 2

Coding

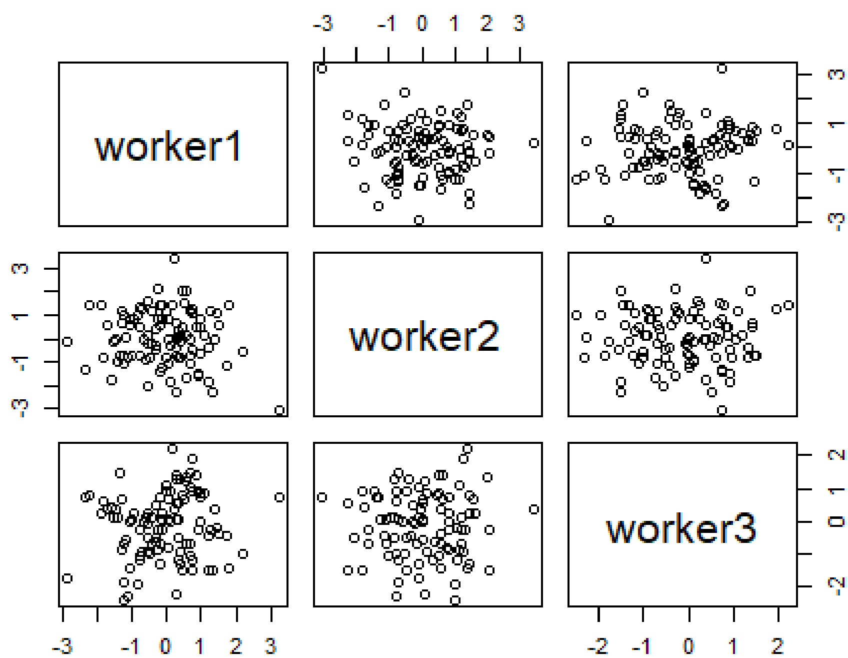
Import data:

```r
df = read.csv('C:/Users/tsong/Downloads/performance_data (1).csv')

head(df,20)
```

Including Plots

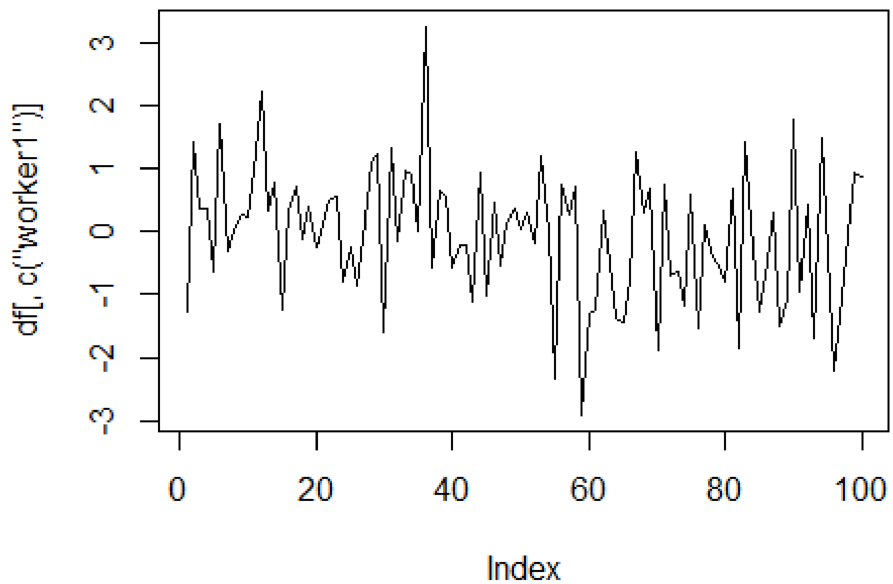
You can also embed plots, for example:

```r
df = read.csv('C:/Users/tsong/Downloads/performance_data (1).csv')
plot(df[,c("worker1","worker2","worker3")])
```

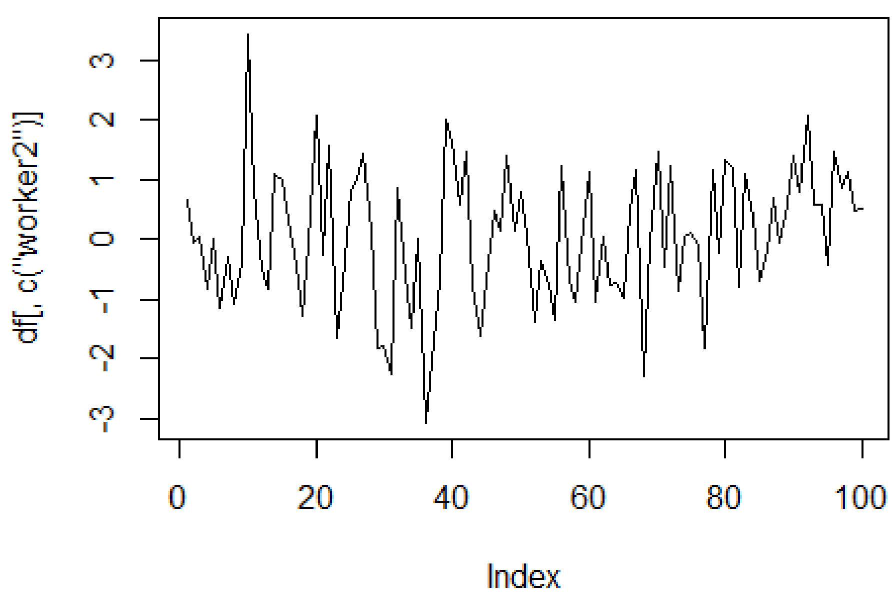


Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

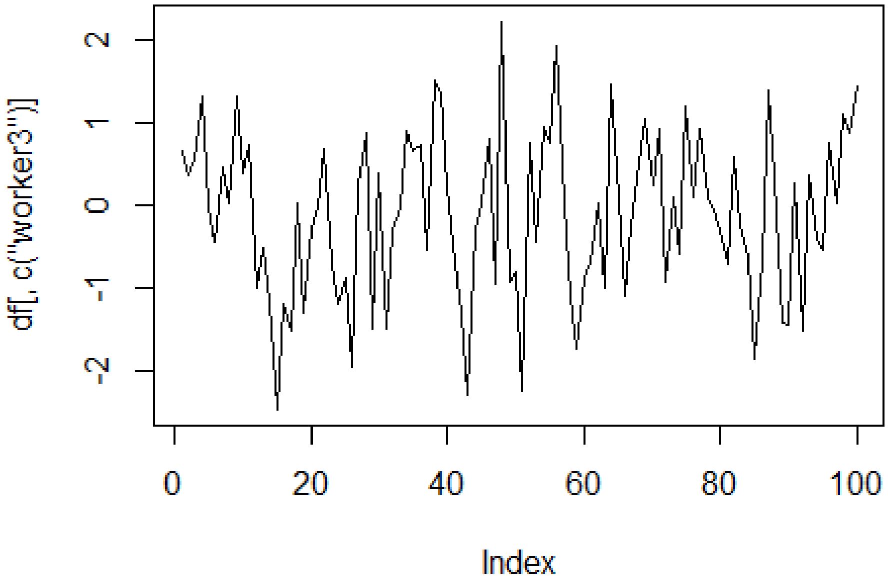
```r
df = read.csv('C:/Users/tsong/Downloads/performance_data (1).csv')
plot(x=df[,c("worker1")], type = 'l')
```




```
plot(x=df[,c("worker2")], type = 'l')
```




```
plot(x=df[,c("worker3")], type = 'l')
```

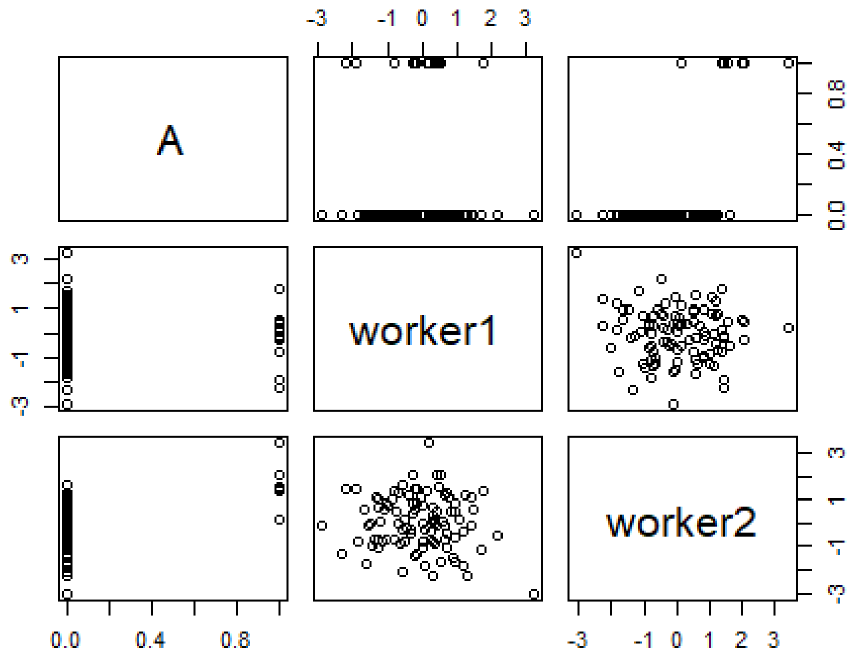


Create two columns to denote interventions on w2 and w3 numerically

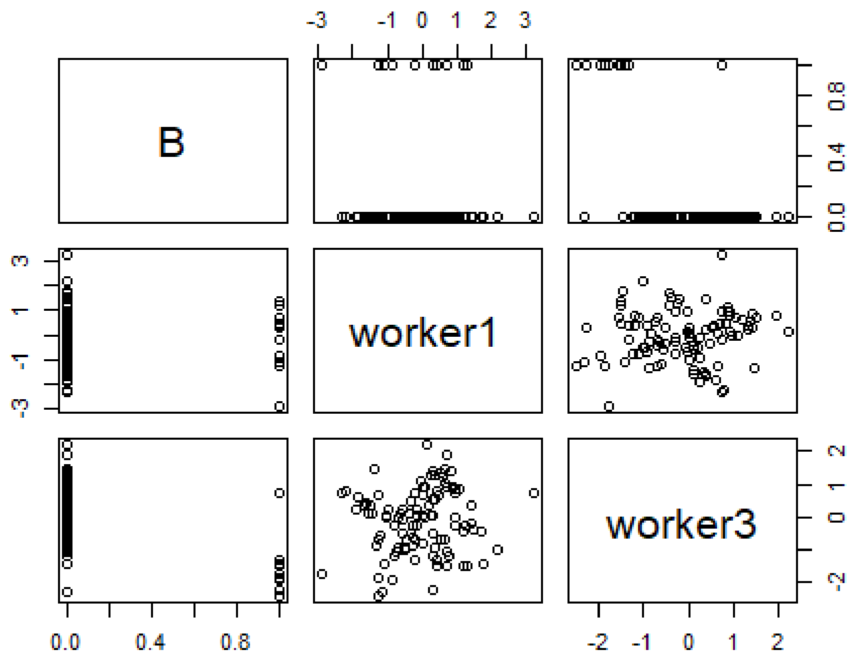
```
df = read.csv('C:/Users/tsong/Downloads/performance_data (1).csv')
df['A'] = ifelse(df["w2_intervention"]=='A',1,0)
df['B'] = ifelse(df["w3_intervention"]=='B',1,0)
```

By plotting the following three variables on scatter plots, we can see that intervention A did have an positive effect on performance as 0 and 1 have different distributions for worker 2

```
df = read.csv('C:/Users/tsong/Downloads/performance_data (1).csv')
df['A'] = ifelse(df["w2_intervention"]=='A',1,0)
df['B'] = ifelse(df["w3_intervention"]=='B',1,0)
plot(df[,c('A','worker1','worker2')])
```



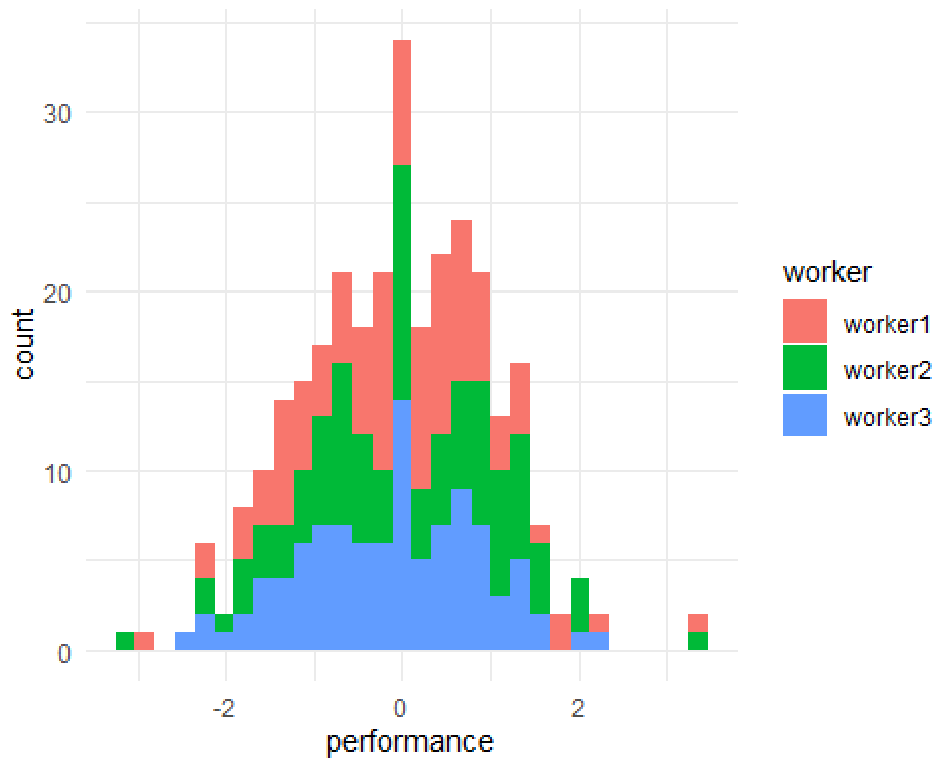

```r
df = read.csv('C:/Users/tsong/Downloads/performance_data (1).csv')
df['A'] = ifelse(df["w2_intervention"]=='A',1,0)
df['B'] = ifelse(df["w3_intervention"]=='B',1,0)
plot(df[,c('B','worker1','worker3')])
```



The intervention on worker has negative effect as the the scatter plot shows the distribution moves downwards

```
library(tidyverse)

## Warning: 程辑包'tidyverse'是用R版本4.3.2 来建造的

## Warning: 程辑包'ggplot2'是用R版本4.3.2 来建造的

## Warning: 程辑包'readr'是用R版本4.3.2 来建造的

## Warning: 程辑包'forcats'是用R版本4.3.2 来建造的

## Warning: 程辑包'lubridate'是用R版本4.3.2 来建造的

## ── Attaching core tidyverse packages ──────────────────────── tidyve
rse 2.0.0 ──
## ✔ dplyr     1.1.3     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.4     ✔ tibble    3.2.1
## ✔ lubridate 1.9.3     ✔ tidyr     1.3.0
## ✔ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_co
nflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

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

data = read.csv('C:/Users/tsong/Downloads/performance_data (1).csv')
df_long = gather(data, key = "worker", value = "performance", worker1,
worker2, worker3)

ggplot(data = df_long, aes(x = performance, fill = worker)) + geom_hist
ogram() + theme_minimal()

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Google Study:

Objective: To assess whether Google employees prefer a salary raise or a bonus

Experimental Design Notation:

R	**O1**	**X**	**O2**
R	**O1**	**X**	**O2**
R	**O1**	**X2**	**O2**
R	**O1**	**X2**	**O2**
R	**O1**		**O2**

1. **R O1 X O2**

- **R:** Randomly assign employees to treatment group
- **O1**: Measure employee satisfaction and compensation expectations (pre-intervention)
- **X:** Implement a salary raise for this group
- **O2:** Measure employee satisfaction and compensation expectations (post-intervention)

2. **R O1 X2 O2**

- **R:** Randomly assign another group of employees to the second treatment group.
- **O1:** Measure employee satisfaction and compensation expectations (pre-intervention).
- **X2**: Implement a bonus for this group.
- **O2:** Measure employee satisfaction and compensation expectations again (post-intervention).

3. **R O1 O2**

- **R:** Randomly assign a third group of employees to the control group
- **O1:** Measure employee satisfaction and compensation expectations (pre-intervention)
- **O2:** Measure employee satisfaction and compensation expectations again, after the other group received a raise (this group does not receive a raise)

In this experimental design, employees are divided into two groups for salary increases and two for bonuses, with each pair of groups receiving different levels of compensation change. The first pair (salary groups) receives two distinct salary increases, and the second pair (bonus groups) receives two different bonus amounts. In the study assessing the impact of salary raises and bonuses on employee satisfaction and preferences, each participant's change in satisfaction is calculated by subtracting their pre-intervention score (O1) from their post-intervention score (O2). This process is repeated for each person in the salary raise group, bonus group, and control group. The average of these changes within each group is then computed to determine the overall impact of each intervention – salary raise and bonus – on employee satisfaction and preferences. Finally, a comparative analysis is conducted by comparing the average changes across the groups

Measurement of employee satisfaction and compensation expectations can be done by:

1. Developing the Survey
    - Satisfaction Questions: Include items that assess various aspects of job satisfaction, like work environment, job role, management, work-life balance, etc.
    - Compensation Expectations: Ask direct questions about preferences for salary raises vs. bonuses. You might use Likert scale questions (e.g., strongly agree to strongly disagree), multiple-choice questions, or ranking systems

Types of Questions:

- Quantitative Questions: These are structured questions that can be statistically analyzed (e.g., "On a scale of 1 to 5, how satisfied are you with your current salary?").
- Qualitative Questions: Open-ended questions to gather more in-depth insights (e.g., "What factors would make you prefer a salary raise over a bonus?")

Potential Threats to Inference and Addressing Them:

1. **Selection Bias:**
- Threat: Non-random selection might lead to unbalanced groups.
- Solution: Random assignment ensures each employee has an equal chance of being in either group, helping balance out pre-existing differences.
- Example: Selecting only engineering department employees for a compensation preference study at a tech company, leading to unbalanced representation compared to including all departments.

2. **Maturation:**
- Threat: Changes over time might affect the outcomes, not the intervention itself.
- Solution: Conduct the study over a short period to minimize the impact of time-related factors.
- Example: Over a year-long study, changes in the economy or personal life events might influence employees' compensation preferences, independent of the study's intervention

3. **Testing Effects:**
- Threat: The pretest itself might influence the posttest responses.
- Solution: Ensure anonymity and confidentiality to reduce the likelihood of response bias due to the pretest
- Example: Employees might become more conscious of their compensation satisfaction after the initial survey, altering their responses in the subsequent survey regardless of actual changes

4. **Instrumentation:**
- Threat: Changes in the way responses are measured between pretest and posttest could affect results.
- Solution: Use the same method and conditions for pretest and posttest measurements.
- Example: Using in-depth interviews for the first survey and an impersonal online questionnaire for the second could lead to differences in response depth and detail, affecting results.

5. **Attrition:**
- Threat: Loss of participants during the study could skew results.
- Solution: Keep the study duration short and engage participants to reduce dropout rates
- Example: If 20 out of 100 employees leave the company during the study, and they have different characteristics or views than those who remain, it could skew the study's final results

6. **Avoiding Cross-Contamination:**

- If the same employees are in more than one group, their experiences from one group could influence their responses in another, leading to biased results

7. **History:**

- Threat: if there's an event that occurs concurrently with the treatment to the test group, the casual inference will be nullified as the effect can be ascribed to one event or another.
- Solution: Ask some questions about their work and life in the survey. Make sure that only the treatment happens with the test group so as to exclude external impact

8. **Regression:**

- Threat: Outliers in observation may be confused with treatment effects.
- Solution: Exclude outliers from the comparison