

# K-Pop Data Analysis

Christine P. Chai  
cpchai21@gmail.com

December 30, 2024

Starting in 2024.

Test citation (Chai, 2024)

## 1 Executive Summary

Write something here

## Disclaimer

The opinions and views expressed in this manuscript are those of the author, and do not necessarily state or reflect those of any institution or government entity.

## 2 R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   : 2.00
## 1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##   Mean  :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
##   Max.  :25.0    Max.    :120.00
```

### 2.1 Read in the Dataset

Idol School (偶像學校)

Motivation: One of the contestants, Snowbaby (蔡瑞雪),<sup>1</sup> is originally from Taiwan.

---

<sup>1</sup>Snowbaby's YouTube channel: <https://www.youtube.com/@snowbaby>

```
library(readxl)
idol_school = read_excel("UNFINISHED_Idol_School_Dataset.xlsx")

columns_to_show = c("Chinese_Name", "English_Name",
                    "Vocal", "Dance", "Physical",
                    "Overall", "Basic_Strength_Ranking")

# UNFINISHED: Prevent columns from truncating!

idol_school[1:10, columns_to_show]
```

```
## # A tibble: 10 x 7
##   Chinese_Name English_Name Vocal Dance Physical Overall Basic_Strength_Ranking
##   <chr>         <chr>         <dbl> <dbl>    <dbl>    <dbl>          <dbl>
## 1 NATTY         NATTY           9.8   8      8.1     8.63           1
## 2 劉怡伶         Tasha           8     9.5    8      8.5            2
## 3 李采映         Lee Chae Yo~    8.5   8.5    7.5     8.17           3
## 4 宋河英         Song Ha You~    8.6   5.9    9.8     8.1            4
## 5 金恩書         Kim Eun Suh     6.3   6.9    10      7.73           5
## 6 金明智         Kim Myong Ji    5.5   7.9    8.2     7.2            6
## 7 張圭悧         Jang Gyuri     7.2   7.1    7       7.1            7
## 8 朴宣           Park Sun       9.5   6.1    5.5     7.03           8
## 9 李悠汀         Lee Yoo Jeo~    5.8   6.2    9       7              9
## 10 金娜妍         Kim Na Yeon     8.3   6      6.4     6.9           10
```

## 2.2 Idol School: Additional Resources

Students who were eliminated from the show:

[https://www.ptt.cc/bbs/fromis\\_9/M.1555819461.A.C73.html](https://www.ptt.cc/bbs/fromis_9/M.1555819461.A.C73.html)

Someone else used random forests to predict the final ranking:

<https://shavid.pixnet.net/blog/post/331691281>

## 2.3 Test for Non-English Characters

CJK = Chinese, Japanese, Korean

Chinese example

RStudio 有辦法打中文嗎？

```
print(" 大家好，很高興能認識你們！")
```

```
## [1] "大家好，很高興能認識你們！"
```

Japanese example

思い出にするにはまだ早すぎる

```
print(" みやわき さくら")
```

```
## [1] "みやわき さくら"
```

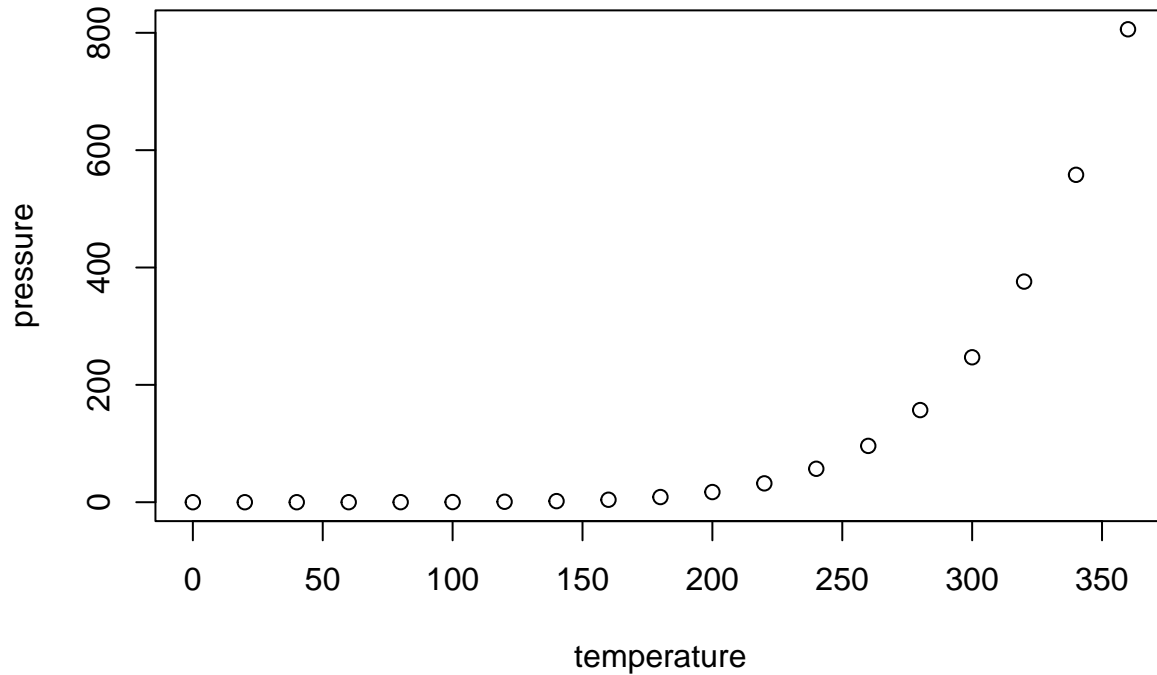
```
print(" 宮脇 咲良")
```

```
## [1] "宮脇 咲良"
```

This template does not support Korean characters yet.

## 2.4 Including Plots

You can also embed plots, for example:



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

## Acknowledgments

Write something here

## References

Chai, C. P. (2024). Statistical analysis of high school and college entrance exam scores in Taiwan with online data. *Preprint on ResearchGate*. <http://dx.doi.org/10.13140/RG.2.2.29468.91520/1>.