原住民族資料分析線上訓練工作坊：R的基礎與應用

第三週

Kacing 廖彥傑
英國艾賽克斯大學博士候選人

數據處理

- 使用readr進行數據導入
- 使用 dplyr 處理關係資料
- 使用 forcats 處理因子 （下周）
- 使用 lubridate 處理日期和時間 （下周）

使用readr數據導入

導入tidyverse

```
library(tidyverse)
```

- `read_csv()` 導入逗號分隔文件
- `read_tsv()` 導入制表符分隔文件
- `read_delim()` 讀取使用任意分隔符的文件

導入資料

```
taitung_county <- read_csv("data/taitung_county.csv")
```

存入資料

```
write_csv(taitung_county, "taitung_county_2.csv")
```

使用readr數據導入

讀取特定列，如`議員`, `年`,與 `性別`

```
select_columns <- read_csv("data/taitung_county.csv",
                           col_select = c("議員", "年", "性別"),
                           show_col_types = FALSE)
```

看前五筆

```
head(select_columns, n = 5)
```

議員	年	性別
江多利	95	1
江多利	95	1
朱連濟	95	0
朱連濟	95	0
江堅壽	95	1

使用 dplyr 處理關係資料

- `%>%`
- 合併連結 (Mutating Joins)
- 篩選
- 集合

使用 dplyr 處理關係資料

讀取資料

```
library(tidyverse)
taitung_county <- read_csv("data/taitung_county.csv",
                           show_col_types = FALSE)
```

鍵 Key

```
taitung_county %>%
  count(議員) %>%
  filter(n > 10) %>%
  head(n=5)
```

議員	n
余忠義	11
嚴惠美	12
宋賢一	57
張國洲	17
張清忠	19

使用 dplyr 處理關係資料

算出哪個年補助款

```
library(tidyverse)
taitung_county %>%
  rename(menber = 議員,
         money = `建議金額(單位: 千元)`,
         year = 年,
         district = `選區(95-100年)`) %>%
  group_by(year) %>%
  mutate(sum_money = sum(money)) %>%
  dplyr::select(year, sum_money) %>%
  dplyr::distinct(year, .keep_all = TRUE)
```

```
#> # A tibble: 6 × 2
#> # Groups:   year [6]
#>    year sum_money
#>   <dbl>     <dbl>
#> 1    95    15414
#> 2    96    28400.
#> 3    97    47599.
#> 4    98   456175.
#> 5    99   233657.
#> 6   100   278316.
```

使用 dplyr 處理關係資料

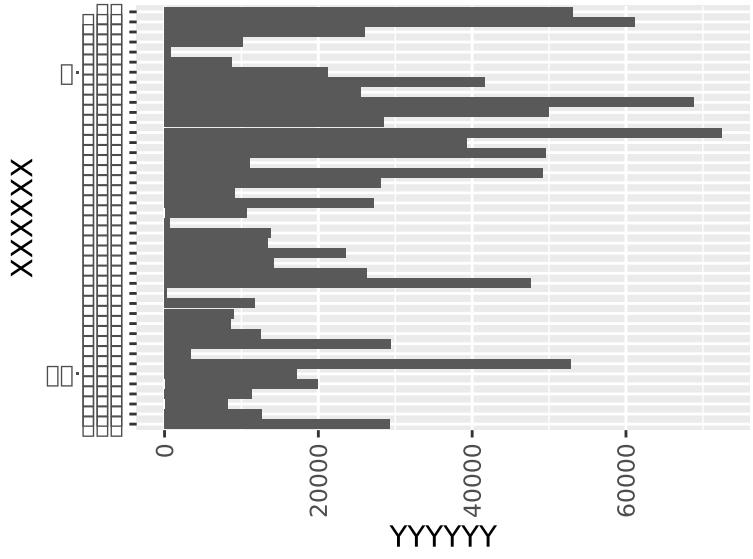
算出每一個議員拿的錢

```
library(tidyverse)
member <- taitung_county %>%
  rename(menber = 議員,
         money = `建議金額(單位: 千元)`,
         year = 年,
         district = `選區(95-100年)`) %>%
  group_by(menber) %>%
  mutate(sum_money = sum(money)) %>%
  dplyr::select(menber, sum_money) %>%
  dplyr::distinct(menber, .keep_all = TRUE)
```

使用 dplyr 處理關係資料

基本視覺化：長條圖

```
member %>%
  ggplot(aes(x=menber, y=sum_money)) +
  geom_bar(stat = "identity") +
  theme(text = element_text(family = "STHeiti")) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  coord_flip() +
  xlab("XXXXXX") +
  ylab("YYYYYY")
```



實作：2019 年文官調查資料

- 文官網路調查資料(合併 2019 年新加樣本與 2018 年追蹤樣本) 中文編碼簿
- 資料連結