Course 1 Section 3.11 - Reading and writing proprietary formats

Jiaying Wu

20/09/2020

```
library(tidyverse)
library(readxl)
library(lubridate)
```

```
# note ".." used becase the my rmd file was not store beside my R  project
path <- "../data/tas_rentals"
xlsx_files <- list.files(path)
xlsx_files
```

```
##  [1] "abs-report-august-2016.xlsx"
##  [2] "abs-report-july-2016.xlsx"
##  [3] "abs-report-june-2016.xlsx"
##  [4] "abs-report-may2016.xlsx"
##  [5] "igate-absreportoct2016-ep.xlsx"
##  [6] "rda-data-report-april-2017.xlsx"
##  [7] "rda-data-report-feb-2017-.xlsx"
##  [8] "rda-data-report-jan2017-ep.xlsx"
##  [9] "rda-data-report-july-2017-.xlsx"
## [10] "rda-data-report-june-2017-.xlsx"
## [11] "rda-data-report-mar-2017-.xlsx"
## [12] "rda-data-report-may-2017.xlsx"
## [13] "rda-datareportdec2016-erik.xlsx"
## [14] "rental-bond-and-rental-data-november-2016.xlsx"
```

```
rentals <- map_dfr(str_c(path, "/", xlsx_files), read_xlsx)
rentals
```

```
## # A tibble: 18,263 x 12
##    `Bond Status` Suburb State Postcode `Bond Amount` `Weekly Rent`
##    <chr>         <chr>  <chr>    <dbl>         <dbl>         <dbl>
##  1 Active        SMITH~ TAS       7330          1000           250
##  2 Active        SMITH~ TAS       7330           800           200
##  3 Active        SMITH~ TAS       7330          1240           310
##  4 Active        STANL~ TAS       7331           480           140
##  5 Active        SMITH~ TAS       7330          1120           280
##  6 Active        LILEAH TAS       7330           960           240
##  7 Active        STANL~ TAS       7331           840           210
##  8 Active        PENGU~ TAS       7316          1200           300
##  9 Active        ST HE~ TAS       7216           980           245
## 10 Active        SCOTT~ TAS       7260           880           220
```

```
## # ... with 18,253 more rows, and 6 more variables: `Bond Lodgement date
## #   (DD/MM/YYYY)` <dttm>, `Bond Activation date (DD/MM/YYYY)` <dttm>, `No of
## #   Bedrooms` <dbl>, `Dwelling/Premises Type` <chr>, `Length of Tenancy (In
## #   Months)` <dbl>, `Street Name` <chr>
```

Wrangling task 1: Rename the columns

```
rentals <- rename(rentals,
                  bond_lodgement_date = `Bond Lodgement date (DD/MM/YYYY)`)
```

Wrangling task 2: Create new columns

```
rentals <- rentals %>%
  mutate(month = month(bond_lodgement_date),
         year = year(bond_lodgement_date))
rentals
```

```
## # A tibble: 18,263 x 14
##    `Bond Status` Suburb State Postcode `Bond Amount` `Weekly Rent`
##    <chr>         <chr>  <chr>    <dbl>         <dbl>         <dbl>
##  1 Active        SMITH~ TAS       7330          1000           250
##  2 Active        SMITH~ TAS       7330           800           200
##  3 Active        SMITH~ TAS       7330          1240           310
##  4 Active        STANL~ TAS       7331           480           140
##  5 Active        SMITH~ TAS       7330          1120           280
##  6 Active        LILEAH TAS       7330           960           240
##  7 Active        STANL~ TAS       7331           840           210
##  8 Active        PENGU~ TAS       7316          1200           300
##  9 Active        ST HE~ TAS       7216           980           245
## 10 Active        SCOTT~ TAS       7260           880           220
## # ... with 18,253 more rows, and 8 more variables: bond_lodgement_date <dttm>,
## #   `Bond Activation date (DD/MM/YYYY)` <dttm>, `No of Bedrooms` <dbl>,
## #   `Dwelling/Premises Type` <chr>, `Length of Tenancy (In Months)` <dbl>,
## #   `Street Name` <chr>, month <dbl>, year <dbl>
```

Wrangling task 3: Remove observations

```
rentals <- rentals %>%
  filter(`No of Bedrooms` >= 1 & `No of Bedrooms` <= 5)
rentals
```

```
## # A tibble: 14,497 x 14
##    `Bond Status` Suburb State Postcode `Bond Amount` `Weekly Rent`
##    <chr>         <chr>  <chr>    <dbl>         <dbl>         <dbl>
##  1 Active        SMITH~ TAS       7330          1000           250
##  2 Active        SMITH~ TAS       7330           800           200
##  3 Active        SMITH~ TAS       7330          1240           310
##  4 Active        STANL~ TAS       7331           480           140
##  5 Active        SMITH~ TAS       7330          1120           280
##  6 Active        LILEAH TAS       7330           960           240
##  7 Active        STANL~ TAS       7331           840           210
```

```
##  8 Active        PENGU~ TAS       7316          1200           300
##  9 Active        ST HE~ TAS       7216           980           245
## 10 Active        GLENO~ TAS       7010          1364           341
## # ... with 14,487 more rows, and 8 more variables: bond_lodgement_date <dttm>,
## #   `Bond Activation date (DD/MM/YYYY)` <dttm>, `No of Bedrooms` <dbl>,
## #   `Dwelling/Premises Type` <chr>, `Length of Tenancy (In Months)` <dbl>,
## #   `Street Name` <chr>, month <dbl>, year <dbl>
```

Wrangling task 4: Compute average weekly rent

```
avg_postcode <- rentals %>%
  group_by(Postcode) %>%
  summarise(Average_Weekly_Rent = mean(`Weekly Rent`))

avg_postcode
```

```
## # A tibble: 110 x 2
##    Postcode Average_Weekly_Rent
##       <dbl>               <dbl>
##  1     7000                343.
##  2     7001                343.
##  3     7002                157.
##  4     7004                395.
##  5     7005                385.
##  6     7007                315
##  7     7008                317.
##  8     7009                285.
##  9     7010                285.
## 10     7011                283.
## # ... with 100 more rows
```

```
avg_month <- rentals %>%
  group_by(month) %>%
  summarise(Average_Weekly_Rent = mean(`Weekly Rent`))

avg_month
```

```
## # A tibble: 11 x 2
##    month Average_Weekly_Rent
##    <dbl>               <dbl>
##  1     1                302.
##  2     2                282.
##  3     3                295.
##  4     4                298.
##  5     5                284.
##  6     6                281.
##  7     7                285.
##  8     8                275.
##  9    10                278.
## 10    11                286.
## 11    12                293.
```

```r
avg_year <- rentals %>%
  group_by(year) %>%
  summarise(Average_Weekly_Rent = mean(`Weekly Rent`))

avg_year
```

```
## # A tibble: 2 x 2
##    year Average_Weekly_Rent
##   <dbl>               <dbl>
## 1  2016                280.
## 2  2017                292.
```

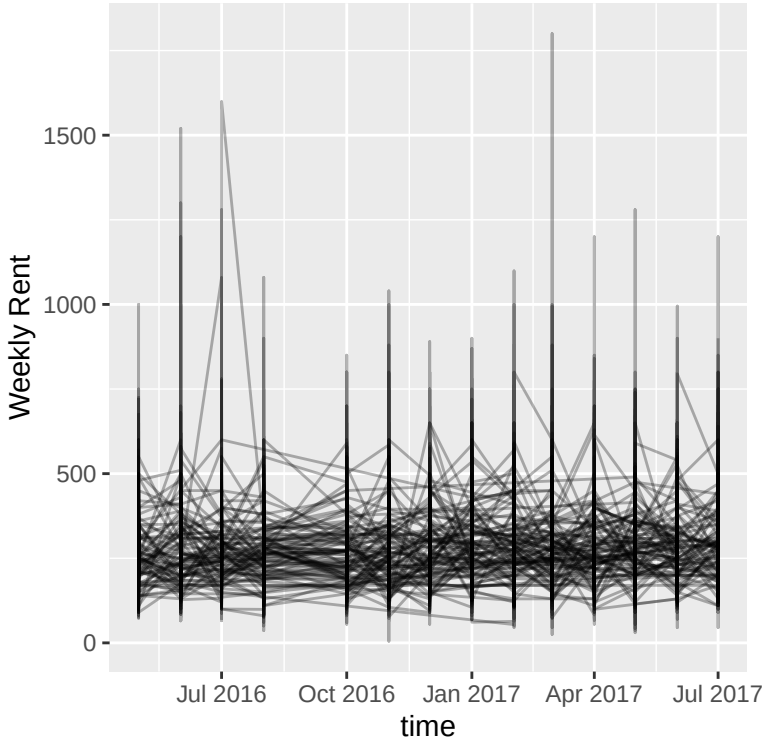
```r
avg_bed <- rentals %>%
  group_by(`No of Bedrooms`) %>%
  summarise(Average_Weekly_Rent = mean(`Weekly Rent`))

avg_bed
```

```
## # A tibble: 5 x 2
##   `No of Bedrooms` Average_Weekly_Rent
##              <dbl>               <dbl>
## 1                1                213.
## 2                2                268.
## 3                3                304.
## 4                4                370.
## 5                5                417.
```

```r
rentals <- rentals %>%
  mutate(time = dmy(paste("01", month, year, sep = "-")))
rentals
```

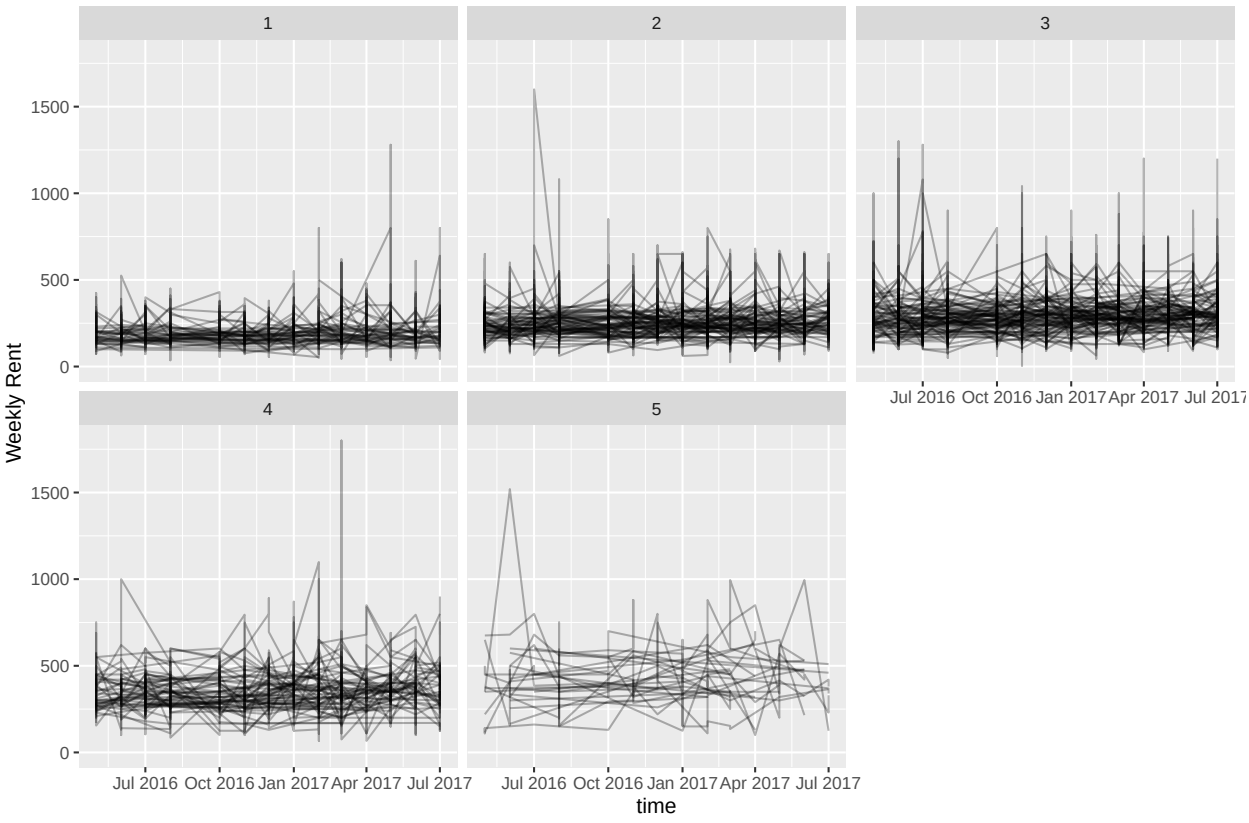
```
## # A tibble: 14,497 x 15
##    `Bond Status` Suburb State Postcode `Bond Amount` `Weekly Rent`
##    <chr>         <chr>  <chr>    <dbl>         <dbl>         <dbl>
##  1 Active        SMITH~ TAS       7330          1000           250
##  2 Active        SMITH~ TAS       7330           800           200
##  3 Active        SMITH~ TAS       7330          1240           310
##  4 Active        STANL~ TAS       7331           480           140
##  5 Active        SMITH~ TAS       7330          1120           280
##  6 Active        LILEAH TAS       7330           960           240
##  7 Active        STANL~ TAS       7331           840           210
##  8 Active        PENGU~ TAS       7316          1200           300
##  9 Active        ST HE~ TAS       7216           980           245
## 10 Active        GLENO~ TAS       7010          1364           341
## # ... with 14,487 more rows, and 9 more variables: bond_lodgement_date <dttm>,
## #   `Bond Activation date (DD/MM/YYYY)` <dttm>, `No of Bedrooms` <dbl>,
## #   `Dwelling/Premises Type` <chr>, `Length of Tenancy (In Months)` <dbl>,
## #   `Street Name` <chr>, month <dbl>, year <dbl>, time <date>
```

```r
rentals %>%
  ggplot(aes(x = time, y = `Weekly Rent`, group = Postcode)) +
  geom_line(alpha = 0.3)
```



Facet your plot

```
rentals %>%
  ggplot(aes(x = time, y = `Weekly Rent`, group = Postcode)) +
  geom_line(alpha = 0.3) +
  facet_wrap(~`No of Bedrooms`)
```


1
2
3
4
5
Weekly Rent
0
500
1000
1500
Jul 2016
Oct 2016
Jan 2017
Apr 2017
Jul 2017
time