Week-4: Code-along

Elise Wong

2023-09-04

II. Code to edit and execute using the Code-along.Rmd file

A. Data Wrangling

1. Loading packages (Slide #16)

```
# Load package tidyverse
library("tidyverse")
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ─
─
## ✔ dplyr     1.1.0     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.3     ✔ tibble    3.1.8
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ─
─
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force a
ll conflicts to become errors
```

2. Loading data-set (Slide #16)

```
# Read data from the hotels.csv file and assign it to a variable named, "hotels"
hotels <- read.csv("hotels.csv")
```

3. List names of the variables in the data-set (Slide #19)

```
# Enter code here
names(hotels)
```

```
##  [1] "hotel"                          "is_canceled"
##  [3] "lead_time"                      "arrival_date_year"
##  [5] "arrival_date_month"             "arrival_date_week_number"
##  [7] "arrival_date_day_of_month"      "stays_in_weekend_nights"
##  [9] "stays_in_week_nights"           "adults"
## [11] "children"                       "babies"
## [13] "meal"                           "country"
## [15] "market_segment"                 "distribution_channel"
## [17] "is_repeated_guest"              "previous_cancellations"
## [19] "previous_bookings_not_canceled" "reserved_room_type"
## [21] "assigned_room_type"             "booking_changes"
## [23] "deposit_type"                   "agent"
## [25] "company"                        "days_in_waiting_list"
```

```
## [27] "customer_type"                  "adr"
## [29] "required_car_parking_spaces"    "total_of_special_requests"
## [31] "reservation_status"             "reservation_status_date"
```

4. Glimpse of contents of the data-set (Slide #20)

```
# Enter code here
glimpse(hotels)
```

```
## Rows: 119,390
## Columns: 32
## $ hotel                           <chr> "Resort Hotel", "Resort Hotel", "Resort…
## $ is_canceled                     <int> 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, …
## $ lead_time                       <int> 342, 737, 7, 13, 14, 14, 0, 9, 85, 75, …
## $ arrival_date_year               <int> 2015, 2015, 2015, 2015, 2015, 2015, 201…
## $ arrival_date_month              <chr> "July", "July", "July", "July", "July",…
## $ arrival_date_week_number        <int> 27, 27, 27, 27, 27, 27, 27, 27, 27, 27,…
## $ arrival_date_day_of_month       <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, …
## $ stays_in_weekend_nights         <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ stays_in_week_nights            <int> 0, 0, 1, 1, 2, 2, 2, 2, 3, 3, 4, 4, 4, …
## $ adults                          <int> 2, 2, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, …
## $ children                        <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ babies                          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ meal                            <chr> "BB", "BB", "BB", "BB", "BB", "BB", "BB…
## $ country                         <chr> "PRT", "PRT", "GBR", "GBR", "GBR", "GBR…
## $ market_segment                  <chr> "Direct", "Direct", "Direct", "Corporat…
## $ distribution_channel            <chr> "Direct", "Direct", "Direct", "Corporat…
## $ is_repeated_guest               <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ previous_cancellations          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ previous_bookings_not_canceled <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ reserved_room_type              <chr> "C", "C", "A", "A", "A", "A", "C", "C",…
## $ assigned_room_type              <chr> "C", "C", "C", "A", "A", "A", "C", "C",…
## $ booking_changes                 <int> 3, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ deposit_type                    <chr> "No Deposit", "No Deposit", "No Deposit…
## $ agent                           <chr> "NULL", "NULL", "NULL", "304", "240", "…
## $ company                         <chr> "NULL", "NULL", "NULL", "NULL", "NULL",
```

```
## $ company                       <chr> "NULL", "NULL", "NULL", "NULL", "NULL",…
## $ days_in_waiting_list          <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ customer_type                 <chr> "Transient", "Transient", "Transient", …
## $ adr                           <dbl> 0.00, 0.00, 75.00, 75.00, 98.00, 98.00,…
## $ required_car_parking_spaces   <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ total_of_special_requests     <int> 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 3, …
## $ reservation_status            <chr> "Check-Out", "Check-Out", "Check-Out", …
## $ reservation_status_date       <chr> "2015-07-01", "2015-07-01", "2015-07-02…
```

B. Choosing rows or columns

5. Select a single column (Slide #24)

```
# Enter code here
select(hotels, lead_time)
```

6. Select multiple columns (Slide #25)

```
# Enter code here
select(hotels, lead_time, agent, market_segment)
```

7. Arrange entries of a column (Slide #28)

```
# Enter code here
arrange(hotels, lead_time)
```

8. Arrange entries of a column in the descending order (Slide #30)

```
# Enter code here
arrange(hotels, desc(lead_time))
```

9. Select columns and arrange the entries of a column (Slide #31)

```
# Enter code here
select(hotels, lead_time)
```

10. Select columns and arrange the entries of a column using the pipe operator (Slide #37)

```
# Enter code here
hotels %>%
  select(lead_time) %>%
  arrange(desc(lead_time))
```

11. Pick rows matching a condition (Slide #44)

```
# Enter code here
hotels %>%
  filter(children >= 1) %>%
    select(hotel, children)
```

12. Pick rows matching multiple conditions (Slide #46)

12. Pick rows matching multiple conditions (Slide #48)

```
# Enter code here
hotels %>%
  filter(children >= 1,hotel == "City Hotel") %>%
    select(hotel, children)
```

13. Non-conditional selection of rows: sequence of indices (Slide #49)

```
# Enter code here
hotels %>% slice(1:5)
```

```
##          hotel is_canceled lead_time arrival_date_year arrival_date_month
## 1 Resort Hotel           0       342              2015               July
## 2 Resort Hotel           0       737              2015               July
## 3 Resort Hotel           0         7              2015               July
## 4 Resort Hotel           0        13              2015               July
## 5 Resort Hotel           0        14              2015               July
##   arrival_date_week_number arrival_date_day_of_month stays_in_weekend_nights
## 1                       27                         1                      0
## 2                       27                         1                      0
## 3                       27                         1                      0
## 4                       27                         1                      0
## 5                       27                         1                      0
##   stays_in_week_nights adults children babies meal country market_segment
## 1                    0      2        0      0   BB     PRT         Direct
## 2                    0      2        0      0   BB     PRT         Direct
## 3                    1      1        0      0   BB     GBR         Direct
## 4                    1      1        0      0   BB     GBR      Corporate
## 5                    2      2        0      0   BB     GBR      Online TA
##   distribution_channel is_repeated_guest previous_cancellations
## 1               Direct                 0                      0
## 2               Direct                 0                      0
## 3               Direct                 0                      0
## 4            Corporate                 0                      0
## 5                TA/TO                 0                      0
##   previous_bookings_not_canceled reserved_room_type assigned_room_type
## 1                              0                 C                  C
## 2                              0                 C                  C
## 3                              0                 A                  C
## 4                              0                 A                  A
## 5                              0                 A                  A
##   booking_changes deposit_type agent company days_in_waiting_list customer_type
## 1               3   No Deposit  NULL    NULL                    0     Transient
## 2               4   No Deposit  NULL    NULL                    0     Transient
## 3               0   No Deposit  NULL    NULL                    0     Transient
## 4               0   No Deposit   304    NULL                    0     Transient
## 5               0   No Deposit   240    NULL                    0     Transient
##   adr required_car_parking_spaces total_of_special_requests reservation_status
## 1   0                           0                         0          Check-Out
## 2   0                           0                         0          Check-Out
## 3  75                           0                         0          Check-Out
## 4  75                           0                         0          Check-Out
## 5  98                           0                         1          Check-Out
##   reservation_status_date
## 1              2015-07-01
## 2              2015-07-01
## 3              2015-07-02
## 4              2015-07-02
## 5              2015-07-03
```

14. Non-conditional selection of rows: non-consecutive/specific indices (Slide #50)

```
# Enter code here
hotels %>%
  slice(1, 3, 5)
```

```
##          hotel is_canceled lead_time arrival_date_year arrival_date_month
## 1 Resort Hotel           0       342              2015               July
## 2 Resort Hotel           0         7              2015               July
## 3 Resort Hotel           0        14              2015               July
##   arrival_date_week_number arrival_date_day_of_month stays_in_weekend_nights
## 1                       27                         1                       0
## 2                       27                         1                       0
## 3                       27                         1                       0
##   stays_in_week_nights adults children babies meal country market_segment
## 1                    0      2        0      0   BB     PRT         Direct
## 2                    1      1        0      0   BB     GBR         Direct
## 3                    2      2        0      0   BB     GBR      Online TA
##   distribution_channel is_repeated_guest previous_cancellations
## 1               Direct                 0                      0
## 2               Direct                 0                      0
## 3                TA/TO                 0                      0
##   previous_bookings_not_canceled reserved_room_type assigned_room_type
## 1                              0                  C                  C
## 2                              0                  A                  C
## 3                              0                  A                  A
##   booking_changes deposit_type agent company days_in_waiting_list customer_type
## 1               3   No Deposit  NULL    NULL                    0     Transient
## 2               0   No Deposit  NULL    NULL                    0     Transient
## 3               0   No Deposit   240    NULL                    0     Transient
##   adr required_car_parking_spaces total_of_special_requests reservation_status
## 1   0                            0                         0          Check-Out
## 2  75                            0                         0          Check-Out
## 3  98                            0                         1          Check-Out
##   reservation_status_date
## 1              2015-07-01
## 2              2015-07-02
## 3              2015-07-03
```

15. Pick unique rows using distinct() (Slide #52)

```
# Enter code here
hotels %>% distinct(hotel)
```

```
##          hotel
## 1 Resort Hotel
## 2   City Hotel
```

C. Creating new columns

16. Creating a single column with mutate() (Slide #56)

```
# Enter code here
hotels %>%
  mutate(little_ones = children + babies) %>%
    select(hotel, little_ones, children, babies)
```

17. Creating multiple columns with mutate() (Slide #58)

17. Creating multiple columns with mutate() (Slide #58)

```
# Enter code here
hotels %>%
    mutate(little_ones = children + babies,
           average_little_ones = mean(little_ones)) %>%
      select(hotel, little_ones, children, babies, average_little_ones)
```

D. More operations with examples

18. count() to get frequencies (Slide #60)

```
# Enter code here
hotels %>%
  count(market_segment)
```

```
##   market_segment     n
## 1       Aviation   237
## 2  Complementary   743
## 3      Corporate  5295
## 4         Direct 12606
## 5         Groups 19811
## 6  Offline TA/TO 24219
## 7      Online TA 56477
## 8      Undefined     2
```

19. count() to get frequencies with sorting of count (Slide #61)

```
# Enter code here
hotels %>%
  count(market_segment, sort = TRUE) # <-- decreasing order of counts
```

```
##   market_segment     n
## 1      Online TA 56477
## 2  Offline TA/TO 24219
## 3         Groups 19811
## 4         Direct 12606
## 5      Corporate  5295
## 6  Complementary   743
## 7       Aviation   237
## 8      Undefined     2
```

20. count() multiple variables (Slide #62)

```
# Enter code here
hotels %>%
  count(hotel, market_segment)
```

```
##           hotel market_segment     n
## 1    City Hotel       Aviation   237
## 2    City Hotel  Complementary   542
## 3    City Hotel      Corporate  2986
## 4    City Hotel         Direct  6093
## 5    City Hotel         Groups 13975
## 6    City Hotel  Offline TA/TO 16747
## 7    City Hotel      Online TA 38748
## 8    City Hotel      Undefined     2
```

```
## 8    City Hotel      Undefined     2
## 9  Resort Hotel  Complementary   201
## 10 Resort Hotel      Corporate  2309
## 11 Resort Hotel         Direct  6513
## 12 Resort Hotel         Groups  5836
## 13 Resort Hotel  Offline TA/TO  7472
## 14 Resort Hotel      Online TA 17729
```

21. summarise() for summary statistics (Slide #63)

```
# Enter code here
hotels %>%
  summarise(mean_adr = mean(adr))
```

```
##   mean_adr
## 1 101.8311
```

22. summarise() by using group_by to find mean (Slide #64)

```
# Enter code here
hotels %>%
  group_by(hotel) %>%
  summarise(mean_adr = mean(adr))
```

```
## # A tibble: 2 × 2
##   hotel        mean_adr
##   <chr>           <dbl>
## 1 City Hotel      105.
## 2 Resort Hotel     95.0
```

23. summarise() by using group_by to get count (Slide #65)

```
# Enter code here
hotels %>%
  group_by(hotel) %>%
  summarise(count = n())
```

```
## # A tibble: 2 × 2
##   hotel        count
##   <chr>        <int>
## 1 City Hotel   79330
## 2 Resort Hotel 40060
```

24. summarise() for multiple summary statistics (Slide #67)

```
# Enter code here
hotels %>%
  summarise(
    min_adr = min(adr),
    mean_adr = mean(adr),
    median_adr = median(adr),
    max_adr = max(adr)
)
```

```
##   min_adr mean_adr median_adr max_adr
## 1   -6.38 101.8311     94.575    5400
```

25. select(), slice() and arrange() (Slide #68)

```
# Enter code here
hotels %>%
  select(hotel, lead_time) %>%
  slice(1:5) %>%
  arrange(lead_time)
```

```
##          hotel lead_time
## 1 Resort Hotel         7
## 2 Resort Hotel        13
## 3 Resort Hotel        14
## 4 Resort Hotel       342
## 5 Resort Hotel       737
```

26. select(), arrange() and slice() (Slide #69)

```
# Enter code here
hotels %>%
  select(hotel, lead_time) %>%
  arrange(lead_time) %>%
  slice(1:5)
```

```
##          hotel lead_time
## 1 Resort Hotel         0
## 2 Resort Hotel         0
## 3 Resort Hotel         0
## 4 Resort Hotel         0
## 5 Resort Hotel         0
```

27. filter() to select rows based on conditions (Slide #73)

```
# Enter code here
hotels %>%
  filter(
    adults == 0,
    children >= 1
    ) %>%
  select(adults, babies, children)
```

28. filter() to select rows based on complicated conditions (Slide #74)

```
# Enter code here
hotels %>%
  filter( adults == 1,
    children >= 1 | babies >=1) %>% # | means OR
  select(adults, babies, children)
```

29. count() and arrange() (Slide #76)

```
# Enter code here
hotels %>%
  count(market_segment) %>%
  arrange(desc(n)) # <-- decreasing order of counts
```

```
##   market_segment     n
## 1      Online TA 56477
## 2  Offline TA/TO 24219
## 3         Groups 19811
## 4         Direct 12606
## 5      Corporate  5295
## 6  Complementary   743
## 7       Aviation   237
## 8      Undefined     2
```

30. mutate(), select() and arrange() (Slide #77)

```
# Enter code here
hotels %>%
  mutate(little_ones = children + babies) %>% # <---
  select(children, babies, little_ones) %>%
  arrange(desc(little_ones))
```

31. mutate(), filter() and select() (Slide #78)

```
# Enter code here
hotels %>%
  mutate(little_ones = children + babies) %>%
  filter(
    little_ones >= 1,
    hotel == "Resort Hotel"
    ) %>%
  select(hotel, little_ones)

hotels %>%
  mutate(little_ones = children + babies) %>%
  filter(
    little_ones >= 1,
    hotel == "City Hotel"
    ) %>%
  select(hotel, little_ones)
```