Exam information and practice exam

STATS 220 Semester One 2023

Exam information How to revise for the exam!

Information about questions Practice exam Past exams

This practice exam is designed to give you an idea of the format of the exam and the kinds of questions you can expect to be given. **The project details referred to in the practice exam are from Semester One 2022.**

Please use Ed Discussion to post questions related to this practice exam, or general revision related questions 😺

Section 1 Section 2 Section 3

This section will contain questions based on general R code knowledge related to vectors and data frames, as well as data manipulations and visualisations (including the grammar of graphics).

Total 22 marks

Data was obtained from the Spotify API for songs on the NZ Top 40 charts the week of the 23rd May 2022.

For reference, the first 10 rows of the data frame `top40_data` are shown below.

```
> top40_data
# A tibble: 40 x 7
   track_name      mode  track_album_relea~ track_duration_~ tempo energy
   <chr>           <chr> <chr>                        <int> <dbl>  <dbl>
 1 As It Was       minor 2022-03-31                  167303 174.   0.731
 2 Late Night Tal~ major 2022-05-20                  177954 115.   0.728
 3 Matilda         major 2022-05-20                  245964 114.   0.294
 4 Music For a Su~ major 2022-05-20                  193813 107.   0.715
 5 First Class     major 2022-04-08                  173947 107.   0.563
 6 About Damn Time minor 2022-04-13                  191822 109.   0.747
 7 go - goddard. ~ minor 2022-03-04                  192514  90.0  0.689
 8 N95             major 2022-05-13                  195950 140.   0.67
 9 Cold Heart - P~ major 2021-08-13                  202735 116.   0.798
10 Heat Waves      major 2020-06-29                  238805  80.9  0.525
# ... with 30 more rows, and 1 more variable: track_album_name <chr>
```

Q1

How many columns/variables are in the data frame `top40_data`?

7

The variable `track_album_release_date` is in column 3. Which function from {lubridate} can be used to convert this variable to `dttm`?

ymd()

What will be the result of running the R code `top40_data$tempo[3]`?

114

What will be the result of running the R code
`top40_data$track_name %>% unique() %>% length()`?

40

How many variables in the data frame `top40_data` are numeric?

3

5 marks

Q2

Suppose the goal is to find the 10 shortest songs in terms of `track_duration`, from those that have tempos greater than 100 beats per minute.

The code below provides the code that a student wrote, but some parts of the code have been replaced with numbers.

```
top40_data %>%
  {1}(track_duration_ms) %>%
  {2}(tempo {3} 100) %>%
  {4}(1 : {5})
```

Use the boxes below to enter the missing function, operator, argument name or value.

{1} arrange

{2} filter

{3} >

{4} slice

{5} 10

5 marks

Q3

How would the code given in Q4 need to be changed in order to find the 20 **longest** songs in terms of in terms of `track_duration`, from those that have tempos greater than 100 beats per minute?

2 marks

Click to reveal answer

Q4

The data frame `day_counts` was used to create a visualisation to compare the weekday that the albums for the songs were released on.

```
> day_counts
# A tibble: 6 x 2
  day_released num_songs
  <chr>            <int>
1 Friday              25
2 Monday               2
3 Sunday               1
4 Thursday             9
5 Tuesday              1
6 Wednesday            2
```

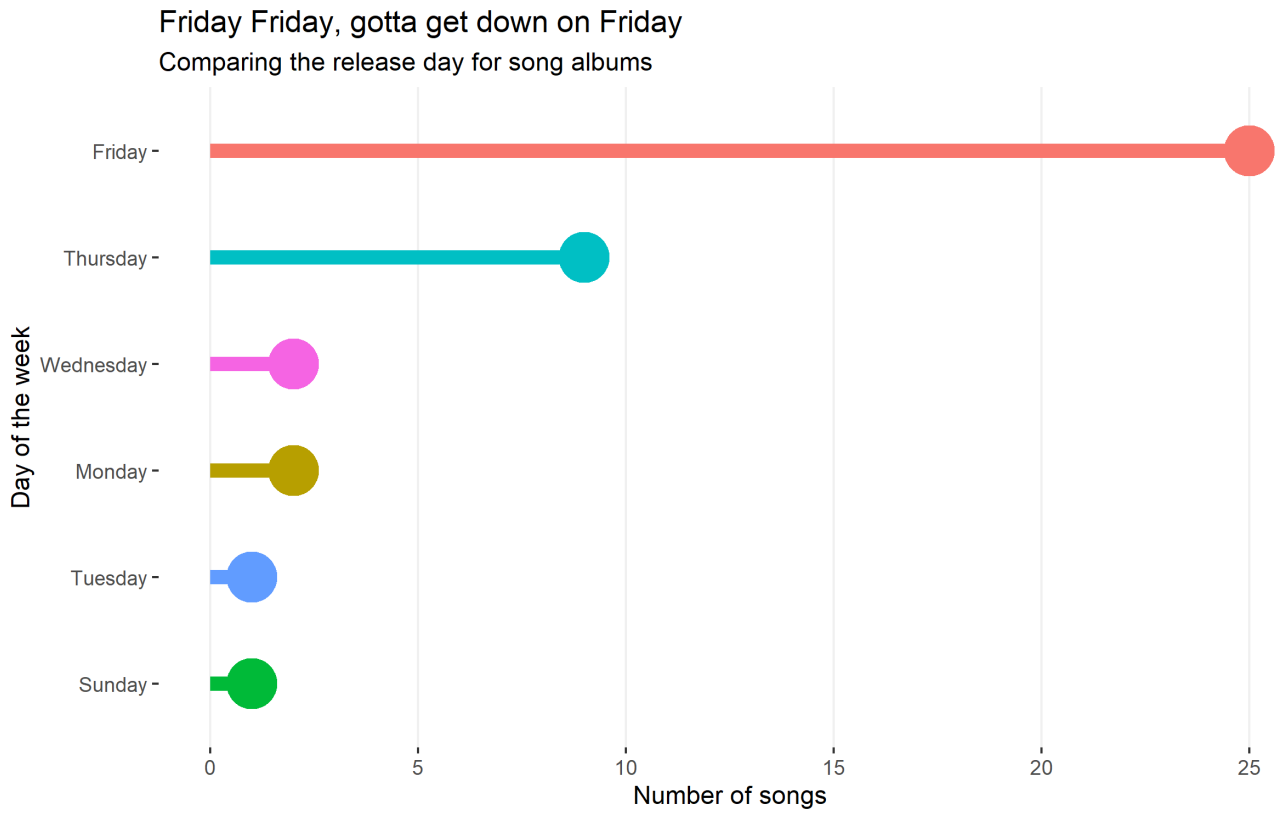
Describe how you could use functions from {dplyr} and {lubridate} to manipulate the data frame `top40_data` to create the data frame `day_counts`.

2 marks

Click to reveal answer

Q5

The visualisation below was created using `day_counts` to compare the weekday that the albums for the songs were released on.



The code below provides the code used to create the visualisation above, but some parts of the code have been replaced with numbers.

```
day_counts %>%
  ggplot({1}({2} = reorder(day_released, num_songs),
             {3} = num_songs,
             colour = day_released)) +
  geom_segment(aes(yend = day_released,
                   xend = 0),
               size = 3) +
  geom_{4}(size = 10) +
  {5}(title = "Friday Friday, gotta get down on Friday",
       subtitle = "Comparing the release day for song albums",
       y = "Day of the week",
       x = "Number of songs") +
  {6}(panel.background = element_rect(fill = "#FFFFFF"),
        panel.grid.major.x = element_line(colour = "#F0F0F0")) +
  guides(colour = "none")
```

Use the boxes below to enter the missing function, operator, argument name or value.

{1} aes

{2} y

{3} x

{4} point

{5} labs

{6} theme

6 marks

Q6

In Project 4, you had to create a visualisation based on music data from a Spotify playlist. **In no more than TWO sentences**, describe what changes you made (or would make using feedback on your project) to improve your visualisation so it communicated a story visually. Refer to the grammar of graphics in your description.

2 marks

Answers will vary, but should cover decisions made such as using titles, labels or annotations, colour, different layers, different geoms, or effective use of animation.