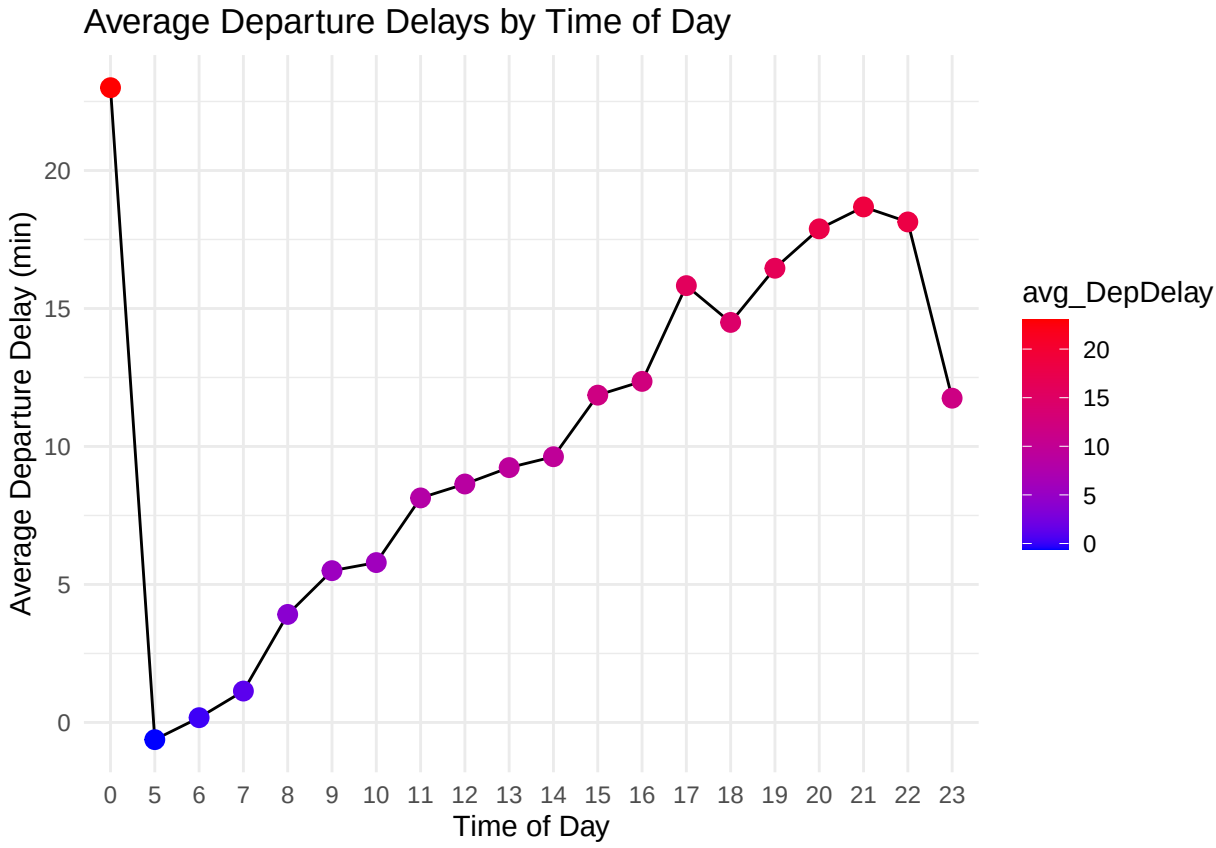
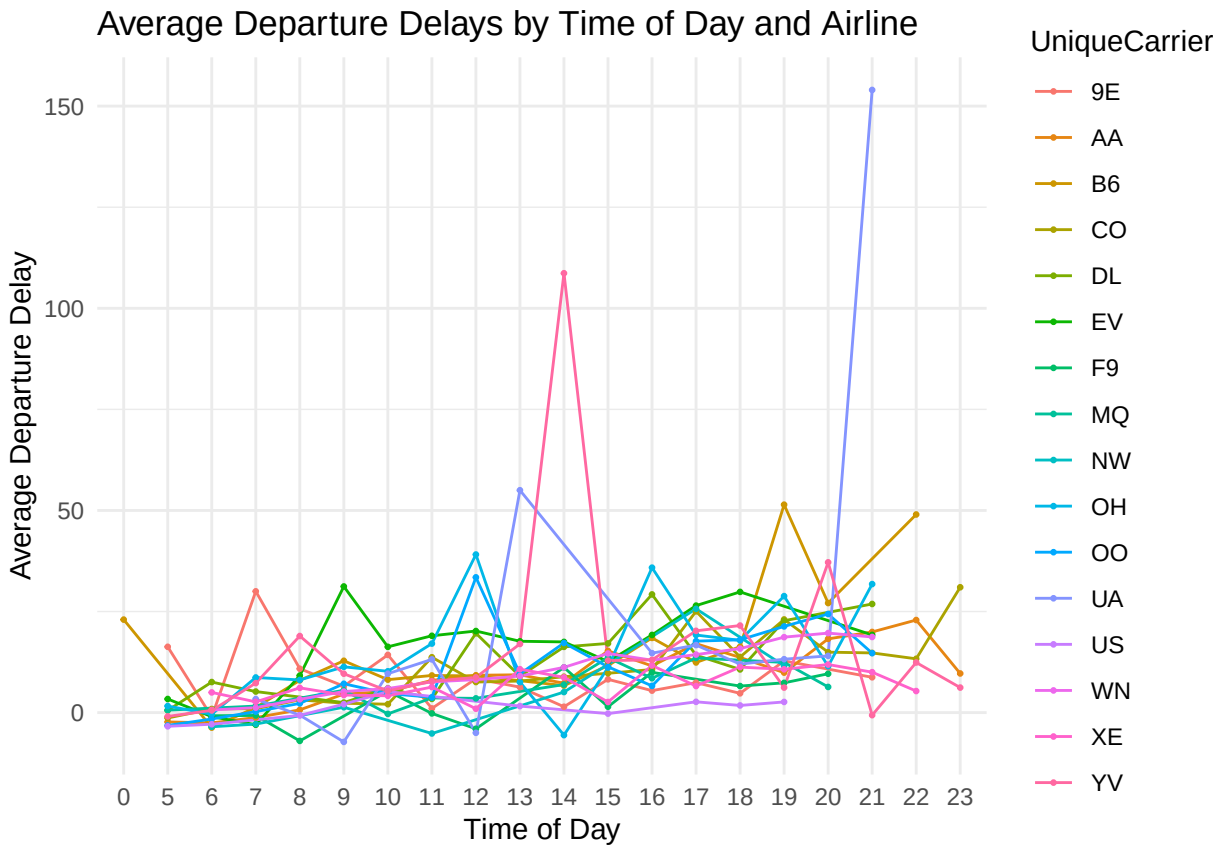
Data Mining HW01

2024-01-31

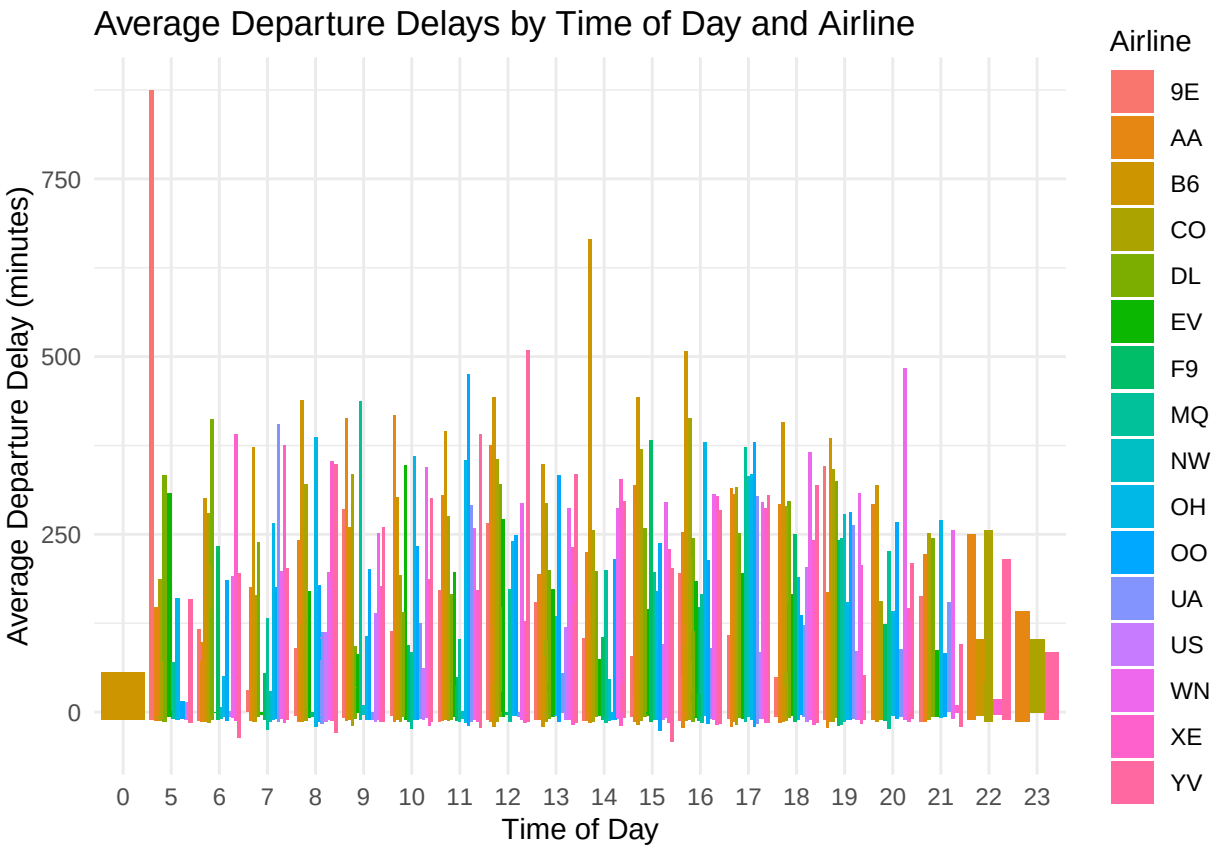
```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
## Rows: 99260 Columns: 29
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (5): UniqueCarrier, TailNum, Origin, Dest, CancellationCode
## dbl (24): Year, Month, DayofMonth, DayOfWeek, DepTime, CRSDepTime, ArrTime, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```



```
## `summarise()` has grouped output by 'CRSDepTimegroup'. You can override using
## the `.groups` argument.
```

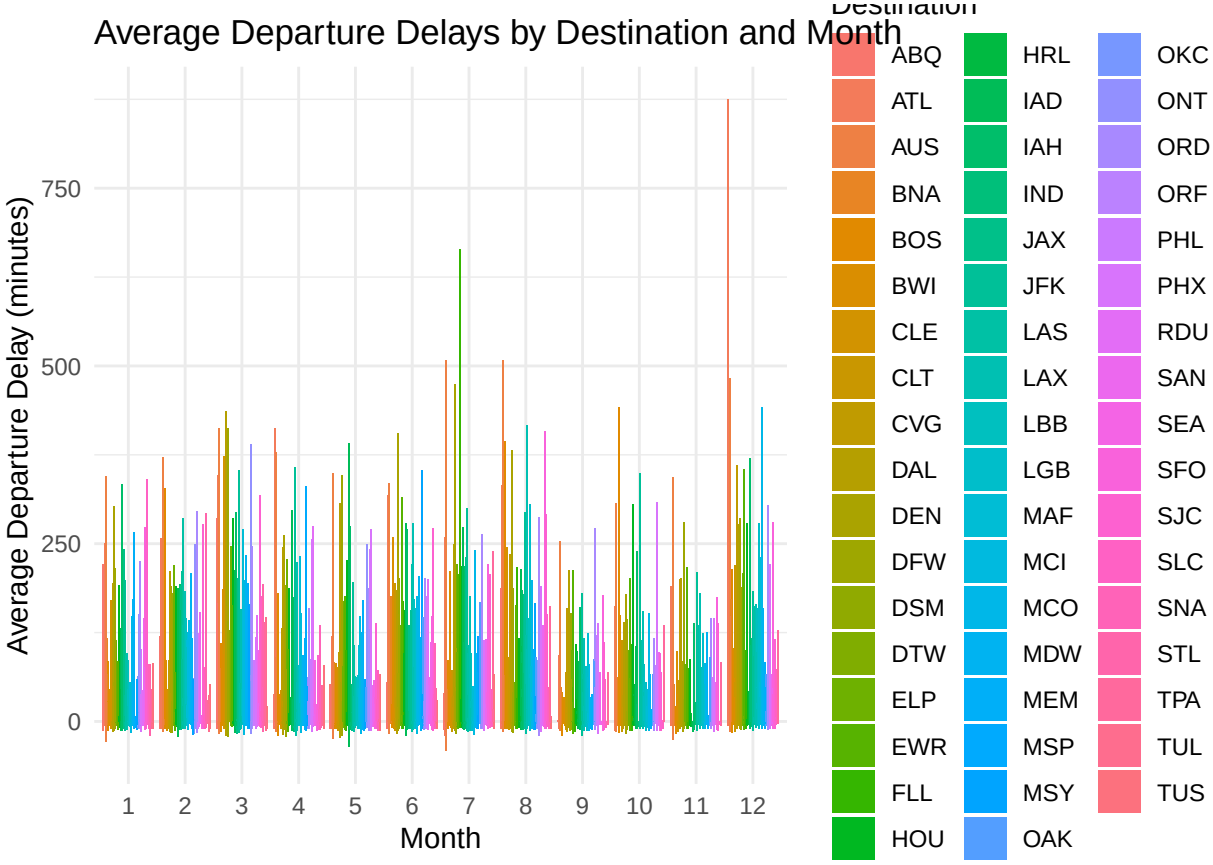



```
## Warning: Removed 1413 rows containing missing values (`geom_bar()`).
```

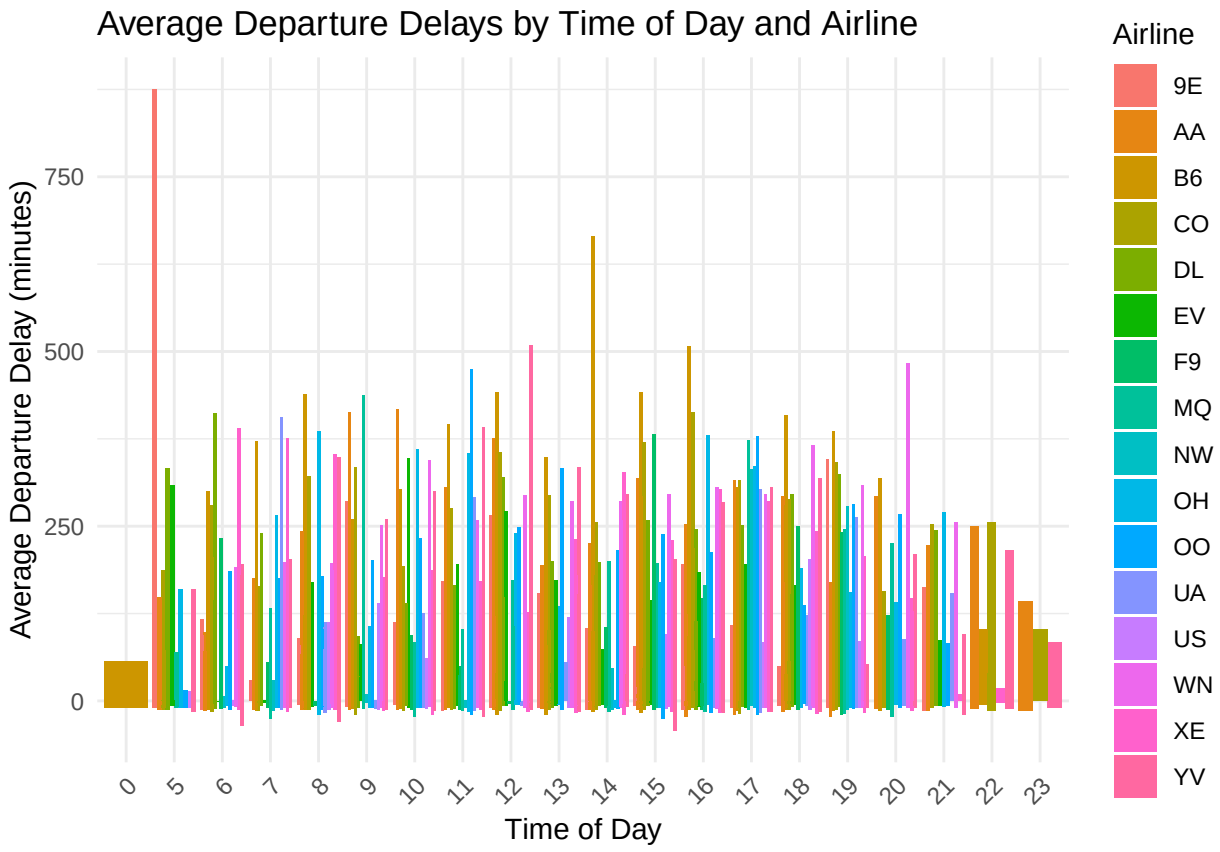


What is the best time of year to fly to minimize delays, and does this change by destination? # Yes, it changes by destination, and avoid summer & winter holidays.

```
## Warning: Removed 1413 rows containing missing values (`geom_bar()`).
```




```
## Warning: Removed 1413 rows containing missing values (`geom_bar()`).
```



Q2

A) What is the 95th percentile of heights for female competitors across all Athletics events (i.e., track and field)? Note that sport is the broad sport (e.g. Athletics) whereas event is the specific event (e.g. 100 meter sprint).

The 95th percentile of heights for female competitors across all Athletics events is 183 cm.

```
## [1] 183
```

B) Which single women's event had the greatest variability in competitor's heights across the entire history of the Olympics, as measured by the standard deviation?

The single women's event with the greatest variability in competitor's heights across the entire history of the Olympics, as measured by the standard deviation, is the Swimming Women's 100 metres Butterfly, with a standard deviation of approximately 8.13 cm.

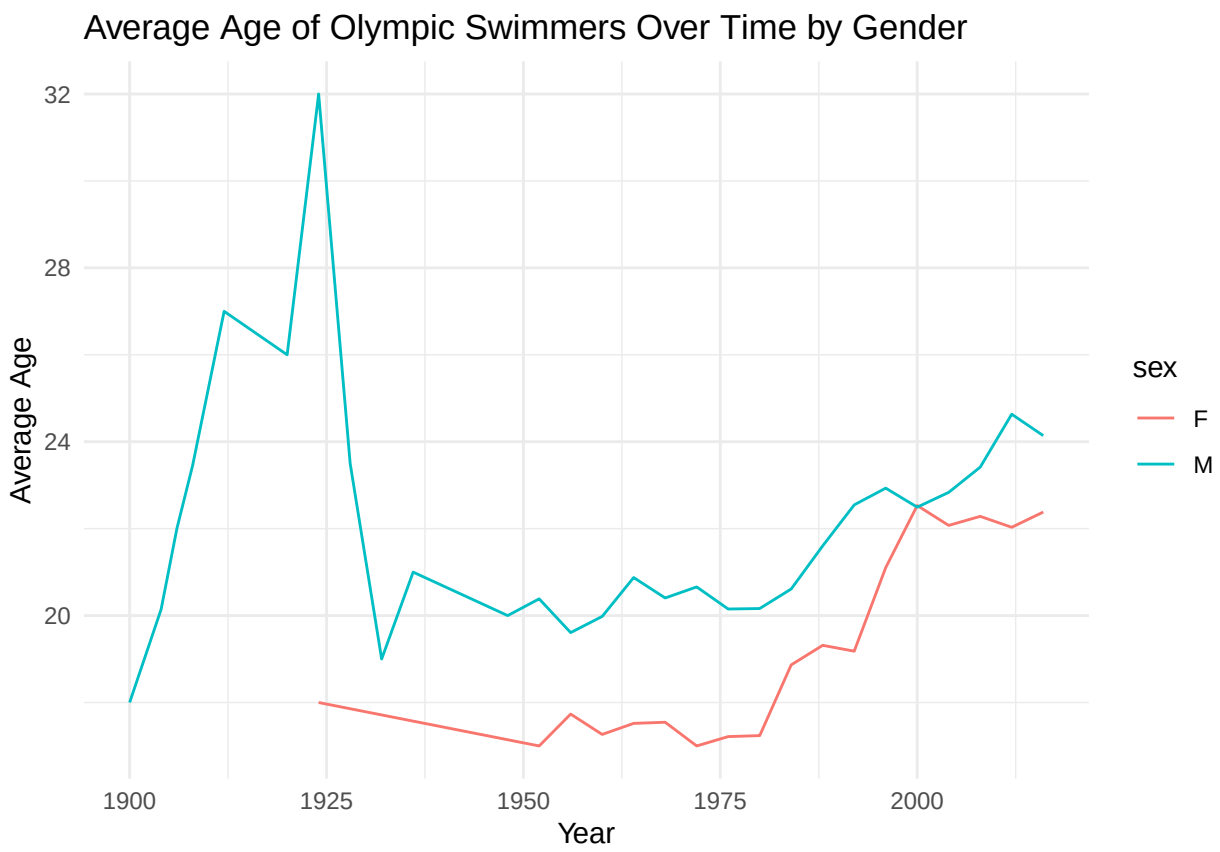
```
## # A tibble: 20 x 2
##    event                                    std_dev_height
##    <chr>                                             <dbl>
##  1 Rowing Women's Coxed Fours                        10.9
##  2 Basketball Women's Basketball                      9.70
##  3 Rowing Women's Coxed Quadruple Sculls              9.25
##  4 Rowing Women's Coxed Eights                        8.74
##  5 Swimming Women's 100 metres Butterfly              8.13
##  6 Volleyball Women's Volleyball                      8.10
##  7 Gymnastics Women's Uneven Bars                     8.02
##  8 Shooting Women's Double Trap                       7.83
##  9 Cycling Women's Keirin                             7.76
## 10 Swimming Women's 400 metres Freestyle              7.62
## 11 Sailing Women's Two Person Dinghy                  7.56
## 12 Shooting Women's Air Pistol, 10 metres             7.26
## 13 Swimming Women's 800 metres Freestyle              7.25
## 14 Cross Country Skiing Women's 20 kilometres         7.23
## 15 Shooting Women's Air Rifle, 10 metres              7.22
## 16 Gymnastics Women's Team All-Around                 7.12
## 17 Athletics Women's Pole Vault                       7.01
## 18 Athletics Women's Triple Jump                      6.96
## 19 Gymnastics Women's Balance Beam                    6.94
## 20 Sailing Women's Skiff                              6.88
```

C) How has the average age of Olympic swimmers changed over time? Does the trend look different for male swimmers relative to female swimmers?

The average age: Both male and female swimmers average age has fluctuated over the years, with a slight increase in recent decades.

Gender difference: There are periods where the average age of female swimmers appears to be slightly lower than that of male swimmers, but the gap between genders seems to have narrowed over time.

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```


Average Age of Olympic Swimmers Over Time by Gender
Average Age
32
28
24
20
1900
1925
1950
1975
2000
Year
sex
F
M