HW data visualization

Pakorn Chiuyoorakkhun
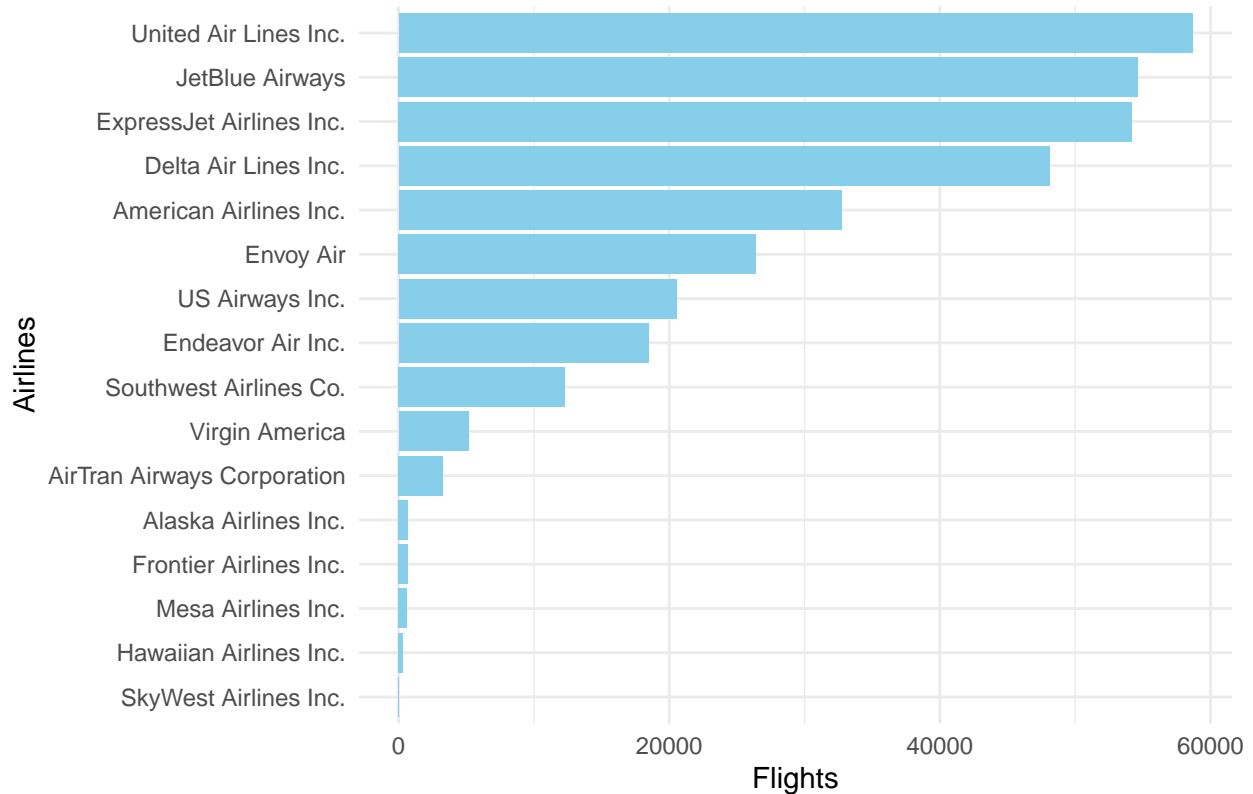
```r
library("tidyverse")
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library("patchwork")
library("nycflights13")
```

```r
flights %>%
  left_join(airlines, by = "carrier") %>%
  group_by(name) %>%
  summarise(flights = n()) %>%
    ggplot( aes(x = reorder(name, + flights), y = flights)) +
      geom_bar(stat = "identity", fill = 'skyblue') +
      coord_flip() +
      labs(
        title = "Ranking of the most flights in 2013",
        x = "Airlines",
        y = "Flights",
          ) +
      theme_minimal()
```

Ranking of the most flights in 2013


```
flights %>%
  left_join(airlines, by = "carrier") %>%
  group_by(name) %>%
  summarise(flights = n()) %>%
  arrange(desc(flights)) %>%
  rename(Airlines = name)
```

```
## # A tibble: 16 x 2
##    Airlines                flights
##    <chr>                     <int>
##  1 United Air Lines Inc.      58665
##  2 JetBlue Airways            54635
##  3 ExpressJet Airlines Inc.   54173
##  4 Delta Air Lines Inc.       48110
##  5 American Airlines Inc.     32729
##  6 Envoy Air                  26397
##  7 US Airways Inc.            20536
##  8 Endeavor Air Inc.          18460
##  9 Southwest Airlines Co.     12275
## 10 Virgin America              5162
## 11 AirTran Airways Corporation 3260
## 12 Alaska Airlines Inc.         714
## 13 Frontier Airlines Inc.       685
## 14 Mesa Airlines Inc.           601
## 15 Hawaiian Airlines Inc.       342
## 16 SkyWest Airlines Inc.         32
```

```
flights %>%
  left_join(airlines, by = "carrier") %>%
  group_by(name) %>%
  summarise(flights = n()) %>%
  arrange(desc(flights)) %>%
  rename(Airlines = name)
```

```
## # A tibble: 16 x 2
##    Airlines                  flights
##    <chr>                       <int>
##  1 United Air Lines Inc.       58665
##  2 JetBlue Airways             54635
##  3 ExpressJet Airlines Inc.    54173
##  4 Delta Air Lines Inc.        48110
##  5 American Airlines Inc.      32729
##  6 Envoy Air                   26397
##  7 US Airways Inc.             20536
##  8 Endeavor Air Inc.           18460
##  9 Southwest Airlines Co.      12275
## 10 Virgin America               5162
## 11 AirTran Airways Corporation  3260
## 12 Alaska Airlines Inc.          714
## 13 Frontier Airlines Inc.        685
## 14 Mesa Airlines Inc.            601
## 15 Hawaiian Airlines Inc.        342
## 16 SkyWest Airlines Inc.          32
```
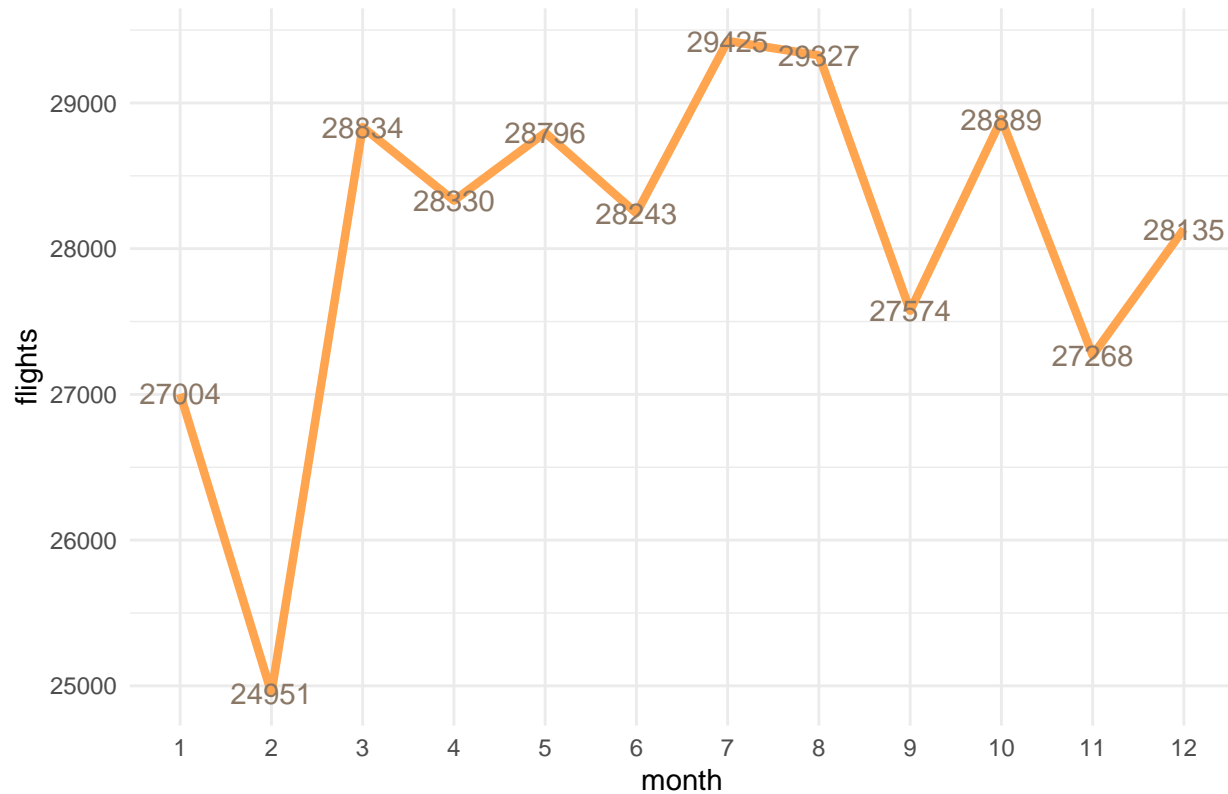
```
flights %>%
  group_by(month) %>%
  summarise(flights = n()) %>%
    ggplot( aes(x = month, y = flights)) +
      geom_line(stat = "identity", color = 'tan1', size = 1.5) +
        scale_x_continuous(breaks = seq(1,12,by=1), minor_breaks = seq(1,12, by = 1)) +
  geom_text(aes(label=flights), color='peachpuff4') +
  labs(
      title = "No of flights in 2013 by month") +

  theme_minimal()
```
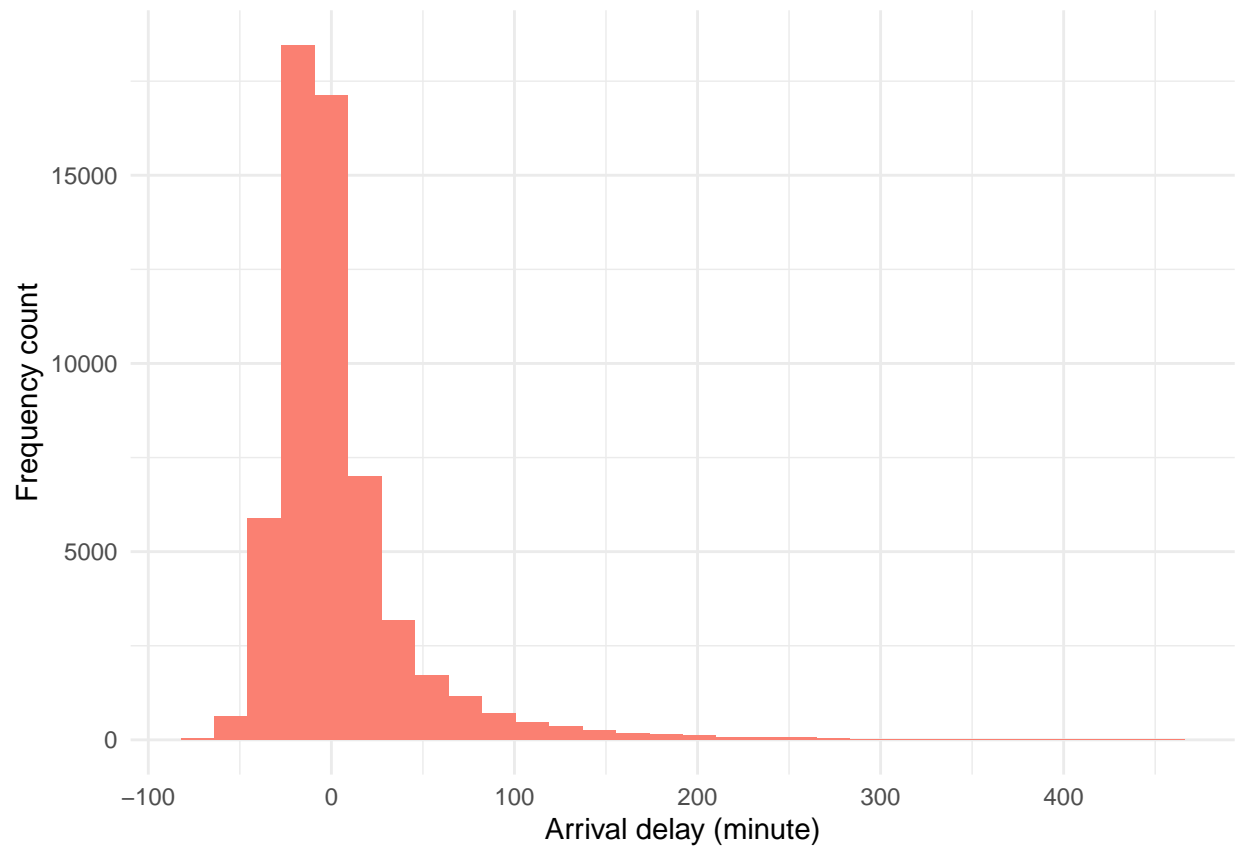
No of flights in 2013 by month
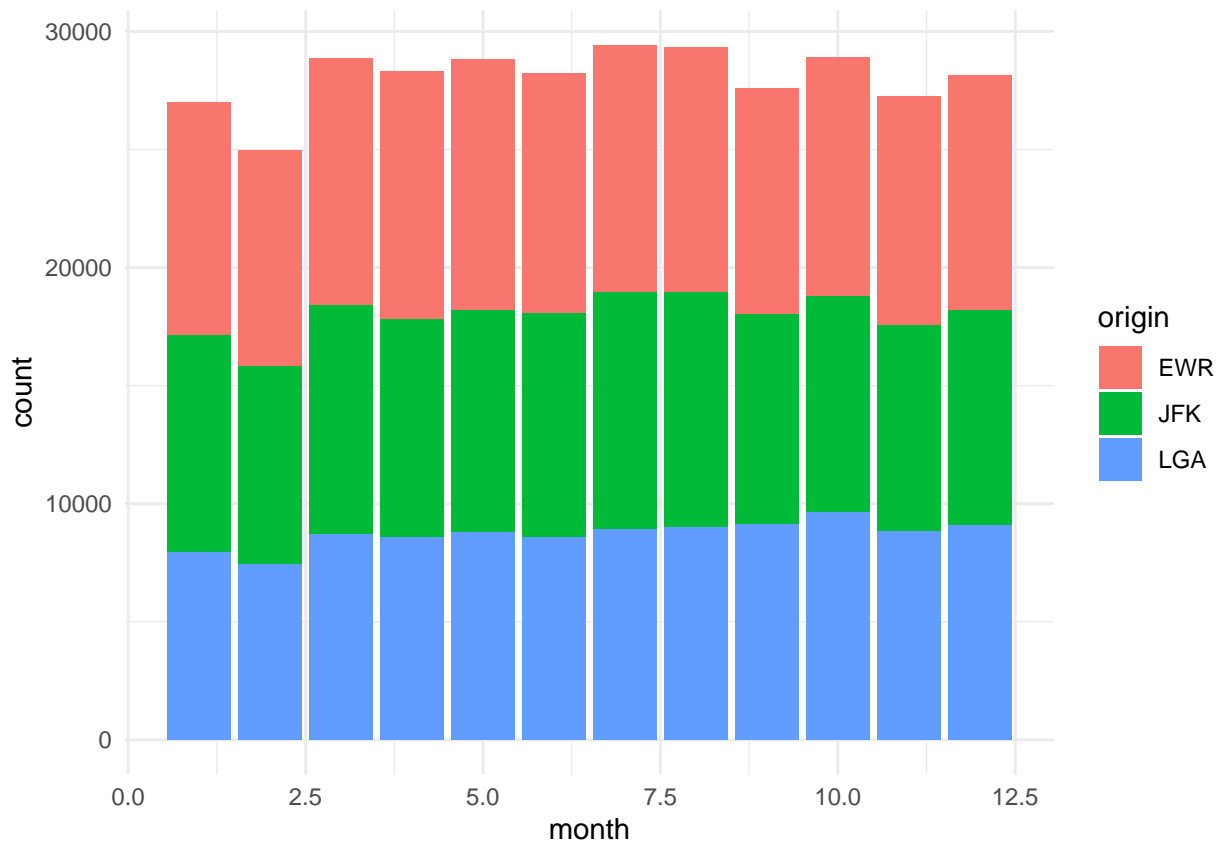


```
flights %>%
  filter(carrier == "UA", !is.na(arr_delay)) %>%
  ggplot(aes(arr_delay)) +
    geom_histogram(fill = 'Salmon') +
  labs(
    x = "Arrival delay (minute)",
    y = "Frequency count"
  ) +
  theme_minimal()
```

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

```
ggplot(flights, aes(month, fill = origin)) +
  geom_bar() +
  theme_minimal()
```

```
flights %>%
  left_join(airlines, by = "carrier") %>%
  group_by(name, month) %>%
  summarise(flights = n()) %>%
  ggplot( aes(x = month, y = flights, color=name)) +
      geom_line(stat = "identity") +
        scale_x_continuous(breaks = seq(1,12,by=1), minor_breaks = seq(1,12, by = 1)) +
        facet_wrap(~name, ncol=4) +
        theme(legend.position="none")
```

`summarise()` has grouped output by 'name'. You can override using the
`.groups` argument.
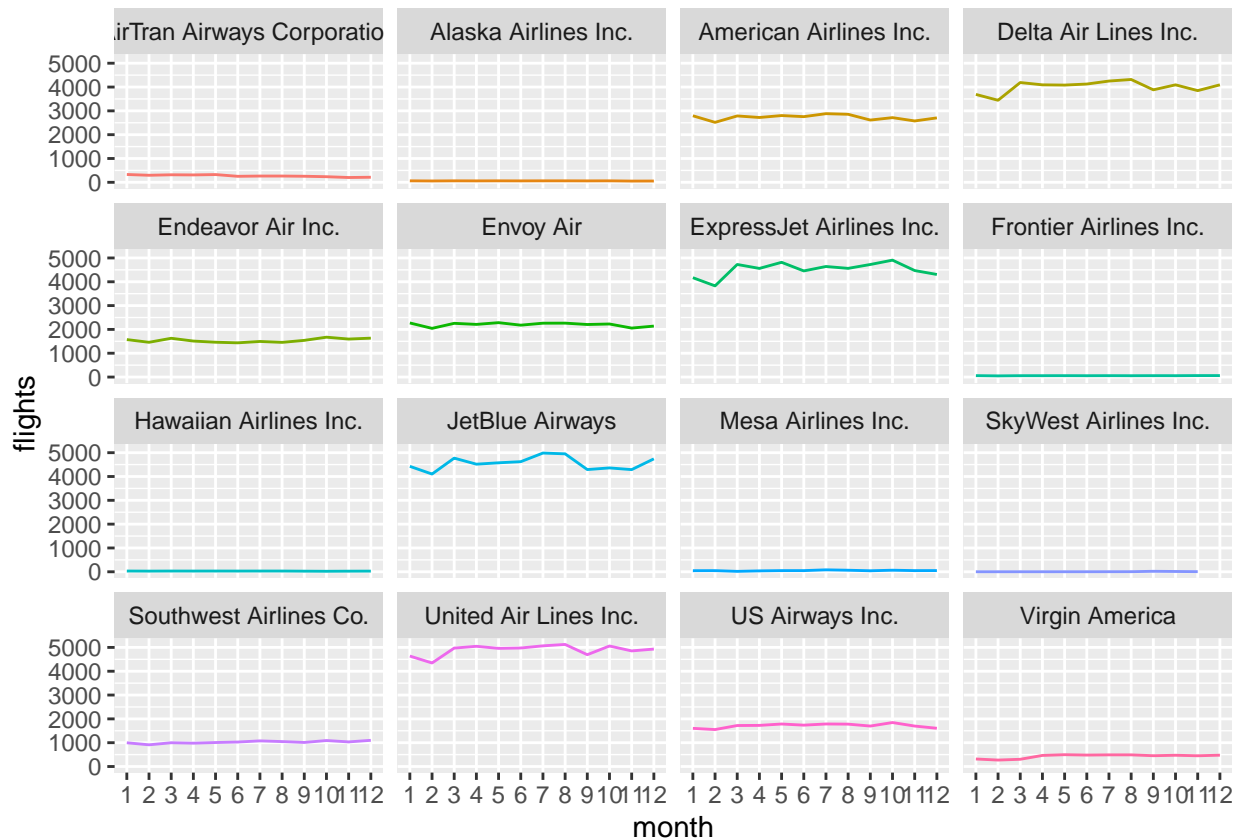
```
flights %>%
  left_join(airlines, by = "carrier") %>%
  group_by(name, month) %>%
  summarise(flights = n()) %>%
  pivot_wider(names_from = month,
              values_from = flights  ) %>%
  rename(Airlines = name)
```

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

## # A tibble: 16 x 13
## # Groups:   Airlines [16]
##    Airlines      `1`   `2`   `3`   `4`   `5`   `6`   `7`   `8`   `9`  `10`  `11`
##    <chr>       <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int>
##  1 AirTran Ai~   328   296   316   311   325   252   263   263   255   236   202
##  2 Alaska Air~    62    56    62    60    62    60    62    62    60    62    52
##  3 American A~  2794  2517  2787  2722  2803  2757  2882  2856  2614  2715  2577
##  4 Delta Air ~  3690  3444  4189  4092  4082  4126  4251  4318  3883  4093  3849
##  5 Endeavor A~  1573  1459  1627  1511  1462  1437  1494  1456  1540  1673  1595
##  6 Envoy Air    2271  2044  2256  2211  2284  2178  2261  2263  2206  2228  2056
##  7 ExpressJet~  4171  3827  4726  4561  4817  4456  4641  4563  4725  4908  4471
##  8 Frontier A~    59    49    57    57    58    55    58    55    58    57    61
##  9 Hawaiian A~    31    28    31    30    31    30    31    31    25    21    25
## 10 JetBlue Ai~  4427  4103  4772  4517  4576  4622  4984  4952  4291  4361  4289
## 11 Mesa Airli~    46    48    18    38    49    49    81    65    42    66    49
## 12 SkyWest Ai~     1    NA    NA    NA    NA     2    NA     4    20    NA     5
## 13 Southwest ~   996   911   998   980  1006  1028  1076  1047  1010  1091  1033
```

```
## 14 United Air~   4637   4346   4971   5047   4960   4975   5066   5124   4694   5060   4854
## 15 US Airways~   1602   1552   1721   1727   1785   1736   1786   1779   1698   1846   1699
## 16 Virgin Ame~    316    271    303    466    496    480    489    489    453    472    451
## # ... with 1 more variable: `12` <int>
```