

```
# NYC Flights
```

```
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
library(tidyverse)
```

Warning message in system("timedatectl", intern = TRUE):

"running command 'timedatectl' had status 1"

Warning message:

"Failed to locate timezone database"

— Attaching packages — tidyverse 1.3.1 —

✓ ggplot2 3.3.5 ✓ purrr 0.3.4

✓ tibble 3.1.5 ✓ stringr 1.4.0

✓ tidyr 1.1.4 ✓ forcats 0.5.1

✓ readr 2.0.2

— Conflicts — tidyverse_conflicts() —

✖ dplyr::filter() masks stats::filter()

✖ purrr::flatten() masks jsonlite::flatten()

✖ dplyr::lag() masks stats::lag()

```
library(readr)
```

```
flights <- read.csv("flights.csv")
```

```
## which carrier has most flight in August in 2013 top 10
```

```
flights %>%
  filter(month == 8, year == 2013) %>%
  count(carrier) %>%
  arrange(desc(n))
```

A data.frame:

```
## Which carrier has the most arrival delay from JFK to MCO and when?
```

```
flights %>%
  select(year, month, day, carrier, origin, dest, arr_delay) %>%
  filter(origin == "JFK", dest == "MCO") %>%
  arrange(desc(arr_delay))
```

A data.frame: 5464 × 7

```
## which carrier has the longest distance ? from which origin to which destination
```

```
flights %>%
  group_by(carrier, origin, dest) %>%
  count(distance) %>%
  arrange(desc(distance)) %>%
  head(5)
```

A grouped_df: 5 × 5

```
## which carrier had an arrival delay of two or more hours
## flew to Houston (IAH or HOU)
## were operated by United, American, or Delta
## Departed in 7, 8, 9
```

```
flights %>%
  select(year, month, day, carrier, origin, dest, arr_delay) %>%
  filter(dest == "IAH",
         arr_delay >= 120,
         carrier %in% c("AA", "UA"),
         month %in% c(7, 8, 9)) %>%
  arrange(desc(arr_delay))
```

A data.frame: 47 x 7

