flights_data_viz

2024-07-28

Show data set

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- 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
```

```
library(nycflights13)
library(glue)
flights
```

```
## # A tibble: 336,776 x 19
##     year month   day dep_time sched_dep_time dep_delay arr_time sched_arr_time
##    <int> <int> <int>    <int>          <int>     <dbl>    <int>          <int>
##  1  2013     1     1      517            515         2      830            819
##  2  2013     1     1      533            529         4      850            830
##  3  2013     1     1      542            540         2      923            850
##  4  2013     1     1      544            545        -1     1004           1022
##  5  2013     1     1      554            600        -6      812            837
##  6  2013     1     1      554            558        -4      740            728
##  7  2013     1     1      555            600        -5      913            854
##  8  2013     1     1      557            600        -3      709            723
##  9  2013     1     1      557            600        -3      838            846
## 10  2013     1     1      558            600        -2      753            745
## # i 336,766 more rows
## # i 11 more variables: arr_delay <dbl>, carrier <chr>, flight <int>,
## #   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>,
## #   hour <dbl>, minute <dbl>, time_hour <dttm>
```

Data Preparation

```
# Create a new data frame with only the necessary columns
flights_routes <- flights %>%
  select(origin, dest)

# Count the frequency of each route and Select the top 10 routes
route_counts <- flights_routes %>%
```

```
  count(origin, dest) %>%
  arrange(desc(n)) %>%
  head(10)

# Create new column "route" showing origin and destination
route <- route_counts %>%
  mutate(route = glue("{origin} to {dest}")) %>%
  arrange(desc(n))

# Convert route to factor for correct ordering
route$route <- factor(route$route, levels = route$route)
```

```
ggplot(route, aes(route, n, fill = n)) +
  geom_col() +
  scale_fill_gradient(low = "lightblue", high = "darkblue") +
  theme(axis.text.x = element_text(size = 5)) +
  labs(title = "Top 10 Popular Routes", x = "Routes", y = "Number of Flights")
```