

# R questions

## 1. Delay Trends by Airline Over the Year

Compute **average arrival delay** by airline and month. Then visualize the trend lines to compare how different airlines perform across the year.

- Functions: `group_by()`, `summarise()`, `mutate()`, `ggplot()` + `geom_line()`

## 2. Weather Impact on Delays

Join `flights` with `weather`. Calculate **average departure delay vs. temperature** buckets (e.g., cut into bins of 5°C). Visualize with a scatter + smoothed regression line.

- Functions: `left_join()`, `mutate(cut())`, `group_by()`, `geom_point()`, `geom_smooth()`

## 3. Flight Distance vs. Air Time by Manufacturer

Join `flights` with `planes`. Plot **distance vs. air\_time**, colored by manufacturer (e.g., Boeing vs. Airbus). Add regression lines for each manufacturer.

- Functions: `left_join()`, `filter()`, `ggplot()` + `geom_point()` + `geom_smooth()`

## 4. Most Reliable Airports

Using `flights` and `airports`, compute **average arrival delay** by destination airport. Plot the **top 15 airports** as a horizontal bar chart (ordered by delay).

- Functions: `left_join()`, `group_by()`, `summarise()`, `arrange(desc())`, `slice_head()`, `geom_col()`, `coord_flip()`

## 5. Distribution of Delays by Airline

For each airline, visualize the **distribution of arrival delays** as a boxplot. This allows comparison of not just averages but variability and outliers.

- Functions: `group_by()`, `ggplot()` + `geom_boxplot()`

## 6. Heatmap of Delays by Day and Hour

Create a **heatmap** of average departure delay by `day` and `hour` across all flights.

- Functions: `group_by()`, `summarise()`, `ggplot()` + `geom_tile()` with `fill = avg_delay`, `scale_fill_gradient()`

## 7. Effect of Wind Speed on Flight Duration

Join with `weather`. Plot **air\_time vs. wind\_speed** to check if high winds shorten or lengthen flights. Use scatterplot + smoothing curve.

- Functions: `left_join()`, `ggplot()` + `geom_point()` + `geom_smooth()`

## 8. Airline Market Share

Compute the **percentage of flights by airline** across the whole dataset. Display as a bar chart sorted from largest to smallest.

- Functions: `count()`, `mutate(prop = n/sum(n))`, `geom_col()`, `reorder()`

## 9. Faceted Trends: Distance vs. Delays by Carrier

Plot **average arrival delay vs. average distance** for each airline. Use facets (`facet_wrap()`) so each airline gets its own panel.

- Functions: `group_by()`, `summarise()`, `ggplot()` + `geom_point()` + `facet_wrap()`

## 10. Flight Volume & Delays Over Time

For each month, compute **number of flights + average delay**. Create a **dual visualization**: bar chart for flights count + line for average delay, using `geom_col()` + `geom_line()`.

- Functions: `group_by()`, `summarise()`, `mutate()`, `ggplot()` + `geom_col()` + `geom_line()`

## 11. Airline Punctuality Ranking

Compute **on-time performance (%)** for each airline (flights arriving  $\leq 15$  min late). Rank carriers and show a **horizontal bar chart** with percentages.

- Functions: `mutate()` , `filter()` , `group_by()` , `summarise()` , `geom_col()` , `coord_flip()`

## 12. Departure Delays by Day of Week

Which days of the week see the worst delays? Summarize average `dep_delay` by weekday and visualize with a **bar chart** ordered Mon–Sun.

- Functions: `mutate(weekday = wday())` , `group_by()` , `summarise()` , `geom_col()`

## 13. Aircraft Size vs. Customer Volume

Do larger planes (by `seats` ) carry proportionally more flights? Join `flights` + `planes` , group by seat category (e.g., small <150, medium 150–250, large >250), and plot **bar chart** of total flights.

- Functions: `left_join()` , `mutate(case_when())` , `group_by()` , `summarise()` , `geom_col()`

## 14. Engine Type and Distance Patterns

Compare average `distance` flown for different `engine` types (from `planes` ). Visualize as **boxplots** to show distribution spread.

- Functions: `left_join()` , `group_by()` , `summarise()` , `geom_boxplot()`

## 15. Manufacturer Reliability Proxy

Which manufacturers' planes (Boeing, Airbus, Embraer, etc.) tend to have longer `air_time` relative to `distance` (efficiency)? Create a **scatterplot** of `distance` vs. `air_time` colored by manufacturer.

- Functions: `left_join()` , `filter()` , `geom_point()` , `geom_smooth()`

## 16. Seasonal Weather Impact on Flight Volume

Use `weather` and `flights` : aggregate **number of flights per month vs. avg temperature** at origin airports. Show with a **dual-axis line chart** (flights vs. temp).

- Functions: `left_join()` , `group_by()` , `summarise()` , `pivot_longer()` , `geom_line()`

## 17. Plane Age and Usage

For each aircraft, compute its **age in 2013** ( `2013 - year` ) and total flights it made. Visualize with a **scatterplot** (age vs. number of flights).

- Functions: `mutate()` , `group_by()` , `summarise()` , `geom_point()`

## 18. Weather and Aircraft Interaction

Do heavier planes (more `seats` ) fly more consistently in bad weather? Join `flights + planes + weather` , group by seat category, and plot **boxplots** of `visib` during flights.

- Functions: `left_join()` , `mutate()` , `group_by()` , `geom_boxplot()`

## 19. Airlines' Fleet Composition

For each airline, show the proportion of flights by **aircraft manufacturer** (Boeing, Airbus, etc.) using a **faceted bar chart** (one facet per airline).

- Functions: `left_join()` , `group_by()` , `count()` , `mutate(prop = n/sum(n))` , `geom_col()` , `facet_wrap()`

## 20. Airport Performance by Distance & Delay

Which airports (by `origin` ) handle the **longest average flight distances** and how does this relate to their **average arrival delay**?

- Preprocessing: join `flights + airports` → group by `origin` → calculate mean `distance` and mean `arr_delay` .
- Visualization: **bubble chart** ( `ggplot2::geom_point` ) with average distance (x), average delay (y), and bubble size = number of flights.

**Functions:** `left_join()` , `group_by()` , `summarize()` , `geom_point()`

## 21. Airport Connection Network

Which airports are the **most connected hubs** (serving the widest variety of destinations)?

- Preprocessing: from `flights` , group by `origin` → count unique `dest` .
- Visualization: **bar chart** sorted descending → top 15 airports by unique destinations.

**Functions:** `group_by()` , `summarize(n_distinct())` , `arrange(desc())` , `geom_col()`

## 22. Airport Elevation Impact on Flight Operations

Does the **elevation** of airports ( `alt` from `airports` ) impact the **average speed** (`distance/air_time`) of flights?

- Preprocessing: join `flights` + `airports` on `origin` → calculate avg flight speed per airport → plot vs airport elevation.
- Visualization: **scatter plot with regression line** ( `geom_point()` + `geom_smooth(method = "lm")` ).