Executive Summary

This project will explore departure delays using the **nycflights13** R package.  
The goals are to analyze how delays vary by **time of day, time of year**, and **weather factors** temperature, wind speed, precipitation, and visibility, using both **exploratory data analysis (EDA)** and **permutation tests.**

High Level Plan

| **Day** | **Tasks** | **Deliverable** | **Responsible** |
| --- | --- | --- | --- |
| **Sat 10/25** | Create shared repo, import nycflights13, confirm variable definitions (late, very\_late). | Project skeleton Rmd, README | **All** |
| **Sun 10/26 – Mon 10/27** | Data prep: filter UA flights, merge with weather, create time and weather bins, QC table. | Cleaned dataset & “Data Methodology” draft | **Member A** |
| **Tue 10/28 – Wed 10/29** | EDA: one plot per factor (6 total), consistent color/theme, clear labels. | Figures + captions | **Member B** |
| **Thu 10/30** | Implement permutation tests; summarize results with p-values & effect sizes. | Result tables | **Member B** |
| **Fri 10/31** | Write Introduction & Results narrative (plain language). | Draft body | **Member C** |
| **Sat 11/1** | Write Discussion, polish writing & formatting, rubric pass. | Near-final draft | **Member C** |
| **Sun 11/2** | Final PDF/HTML knit, check reproducibility, upload submission. | Final report | **All** |

Responsibilities

**Member A – Data & Methodology Lead**

* Filter carrier == "UA", remove NAs in dep\_delay.
* Merge flights with weather by origin and time\_hour.
* Engineer:
  + time\_of\_day bins (e.g., Early 5–9, Midday 10–14, Afternoon 15–18, Evening 19–23, Overnight 0–4)
  + season or month
  + Quantile bins for temperature/wind; binary flags for precipitation/visibility
* Output clean CSV and summary table.
* Write Data Methodology section.

**Member B – Analysis & Visualization Lead**

* Produce consistent ggplots (color palette, titles, captions).
* Conduct permutation tests (use infer or manual shuffle):
  + Binary: difference in means/proportions
  + Multi-group: variance or max difference
* Document each test’s purpose, result, and plain-language meaning.
* Populate Results table

**Member C – Writing & Presentation Lead**

* Write Introduction
* Document Data Methodology. Describe how the data was prepared, cleaned, merged, etc.
* Write Results narrative.
* Write Discussion.
* Final proofreading.
* Compile final Executive Summary and finalize R Markdown output.

Output

The report should address the relationship between departure delays and each of the following:

1. Time of day
2. Time of year
3. Temperature
4. Wind speed
5. Precipitation
6. Visibility

For each of these factors, we should decide whether to use the original dep\_delay variable, or the late and very\_late variables created for the first homework assignment (for some questions, you may only be able to conduct a permutation test (based on what we've learned so far) if you use the late and very\_late variables). If there are circumstances where you feel it is appropriate to create new variables based on the other factors being analyzed, you may do so but be sure to clearly indicate that you have done that.

Repo:

<https://github.com/gpskentzos/DATA5300-Project1>