

Quiz 2 — Sample 2

DATA 412/612

Cumulative through October 28 (weeks 1–9). The new material is `tidyr`, mutating and filtering joins, databases, and strings.

Write your answers in an R script. You may use R and RStudio. You may **not** use the internet or your notes. Load the tidyverse and `nycflights13`.

1. Compute both an inner join and a left join of `visits` to `branches` by `branch`. In one sentence each, say which rows of `visits` the inner join drops and which columns the left join still contains for an unmatched branch.

```
branches <- tibble(
  branch = c("Shaw", "Rosslyn"),
  seats = c(80, 60)
)
visits <- tibble(
  patron = c("Ari", "Bea", "Cal"),
  branch = c("Shaw", "Petworth", "Rosslyn")
)
```

2. From `flights`, keep only the January rows, and from those keep the distinct `carrier` values. Then use a filtering join so that `airlines` contains only carriers that flew in January. How many airlines remain?
3. From each string below, extract the five-digit ZIP code when the string is two letters, a comma, one space, and then five digits. Otherwise the result for that string should be missing.

```
c("MD 21228", "WA, 76731", "VA: 34124", "FL, 21225", "ca, 45691")
```

4. Split `code` into `carrier` and `flight`. `flight` should be an integer.

```
codes <- tibble(code = c("AA-101", "DL-303", "UA-22"))
```

5. Reshape `wide` to one row per carrier and month, with a numeric `month` and a `delays` column. Drop missing delay counts.

```
wide <- tibble(
  carrier = c("AA", "DL"),
  m1 = c(4, NA),
  m2 = c(1, 3)
)
```