Lab One- Josh Goldhaber

2025-09-09

```r
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   3.5.2     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- 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
```

```r
library(ggplot2)
library(skimr)
```

```r
pizza <- readr::read_csv("pizza.csv")
```

```
## New names:
## Rows: 1209 Columns: 17
## -- Column specification
## ------------------------------------------------------ Delimiter: "," chr
## (4): area, operator, driver, quality dbl (10): ...1, index, week, weekday,
## count, price, delivery_min, temperatu... lgl (2): rabate, wrongpizza date (1):
## date
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

3. Create a descriptive statistics table. Feel free to use any package or method you feel works best.

```r
skim_without_charts(pizza)
```

Table 1: Data summary

Name	pizza
Number of rows	1209
Number of columns	17
Column type frequency:	
character	4

Date	1
logical	2
numeric	10

Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
area	10	0.99	5	11	0	3	0
operator	8	0.99	5	7	0	3	0
driver	5	1.00	6	9	0	7	0
quality	201	0.83	3	6	0	3	0

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
date	32	0.97	2014-03-01	2014-03-31	2014-03-16	31

Variable type: logical

skim_variable	n_missing	complete_rate	mean	count
rabate	12	0.99	0.50	FAL: 601, TRU: 596
wrongpizza	4	1.00	0.07	FAL: 1122, TRU: 83

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
...1	0	1.00	605.00	349.15	1.00	303.00	605.00	907.00	1209.00
index	0	1.00	605.00	349.15	1.00	303.00	605.00	907.00	1209.00
week	32	0.97	11.40	1.33	9.00	10.00	11.00	13.00	14.00
weekday	32	0.97	4.44	2.02	1.00	3.00	5.00	6.00	7.00
count	12	0.99	3.44	1.56	1.00	2.00	3.00	4.00	8.00
price	12	0.99	48.73	21.63	8.79	30.98	46.76	63.18	134.33
delivery_min	0	1.00	25.65	10.84	8.80	17.40	24.40	32.50	65.60
temperature	39	0.97	47.94	9.94	19.30	42.23	50.00	55.30	64.80
wine_ordered	12	0.99	0.16	0.36	0.00	0.00	0.00	0.00	1.00
wine_delivered	12	0.99	0.14	0.34	0.00	0.00	0.00	0.00	1.00

4. Produce a histogram on the temperature of pizza upon delivery. Is temperature skewed, if so, how?

```
pizza |>
  ggplot( aes(x = temperature, )) +
    geom_histogram(binwidth = 3) +
    labs(
      x = "Pizza Temp",
```
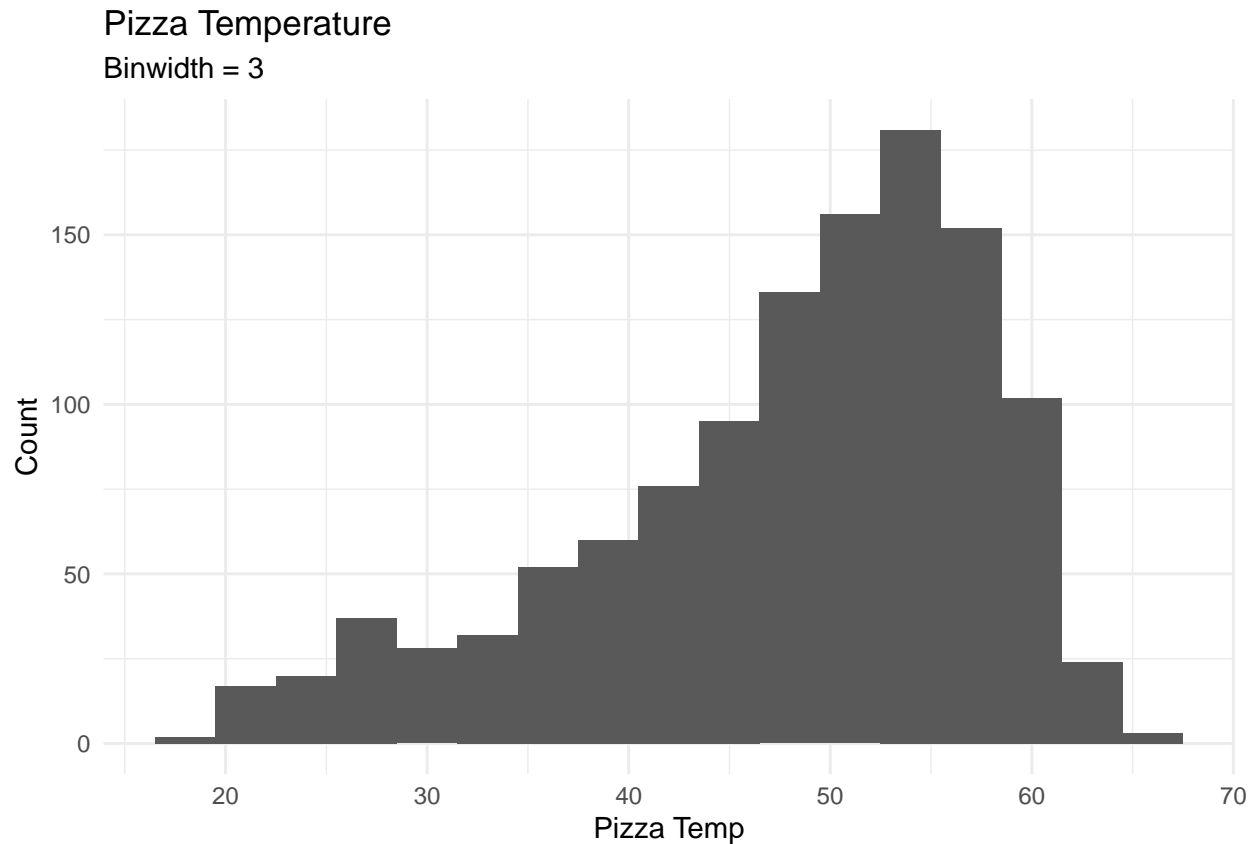
```
    y = "Count",
    title = "Pizza Temperature",
    subtitle = "Binwidth = 3"
  ) +theme_minimal()
```

```
## Warning: Removed 39 rows containing non-finite outside the scale range
## ('stat_bin()').
```


Pizza Temperature
Binwidth = 3

temperature is skewed left.

5. Produce a barplot comparing the average time for delivery by branch location.

```
pizza |>
  group_by(area) |>
  summarise(meanDeliveryTime = mean(delivery_min, na.rm = TRUE)) |>
  ggplot(aes(x = area, y = meanDeliveryTime, fill = area)) +
  geom_col() +
  labs(
    x = "Branch",
    y = "Average Delivery Time (mins)",
    title = "Average Delivery Time by Branch"
  ) +
  theme_minimal()
```

Average Delivery Time by Branch
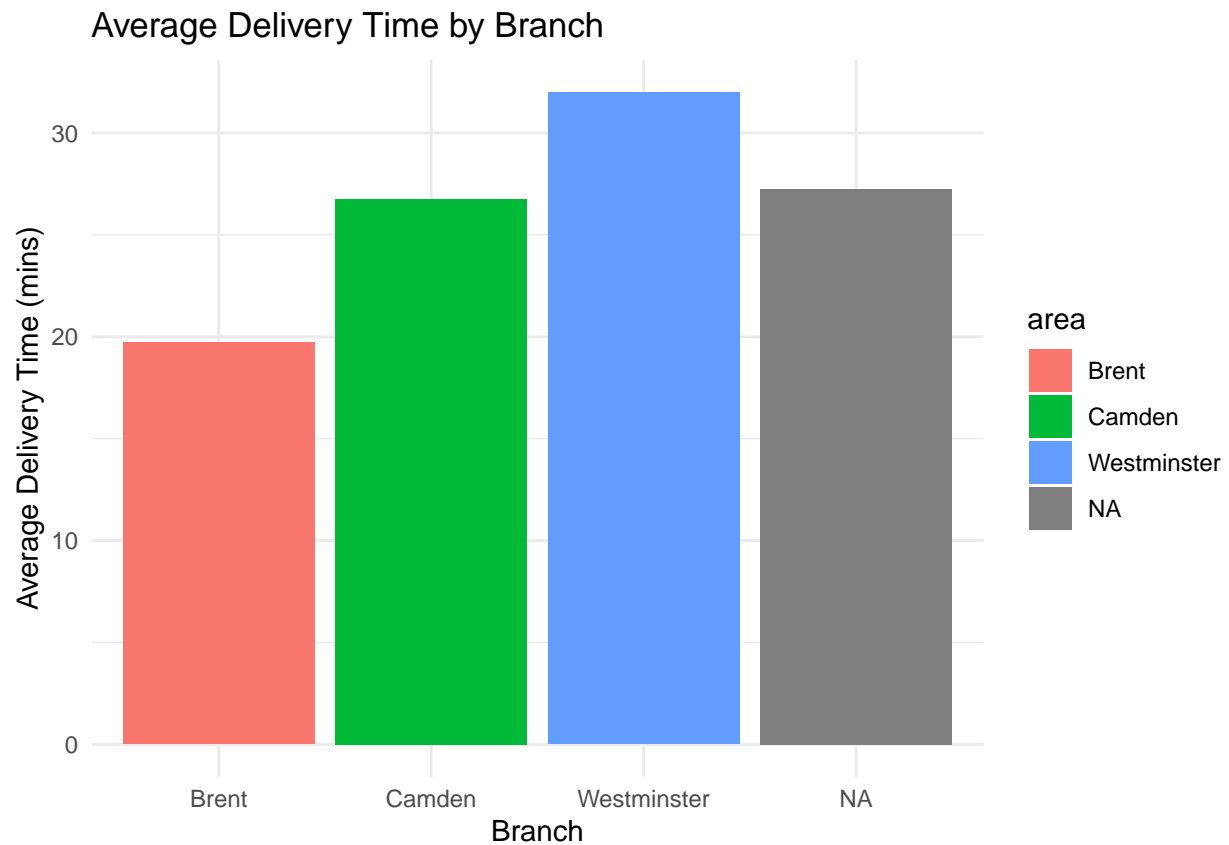
6. Create a boxplot to for delivery time comparing the times across the three branches.

```r
pizza |>
  group_by(area)|>
  ggplot(aes(x = area, y = delivery_min, fill = area))+
  geom_boxplot() +
  labs(
    x = "Branch",
    y = "Delivery Time",
    title = "Delivery Time per Branch",
    fill = "Areas"
  )
```

Delivery Time per Branch