

Data Quality Report: Credit Card Payment

Bryce Charette

Clean data

```
# Adjust column names
colnames(dat) <- c("rec_num", "card_num", "date", "merch_num", "merch_desc",
                  "merch_state", "merch_zip", "trans_type", "amount", "fraud")

# Adjust data types
dat$rec_num <- as.character(dat$rec_num)
dat$card_num <- as.character(dat$card_num)

# Save data
# saveRDS(dat, "card_payments.rds")
```

Statistics & Distributions

rec_num

- **Interpretation:** Unique ID for each record in the data set.
- **Type:** ID
- **Unique Values:** 95,484
- **Missing Values:** 0 (0.00%)
- **Mean:** NA
- **Standard Deviation:** NA
- **Distribution:** NA

card_num

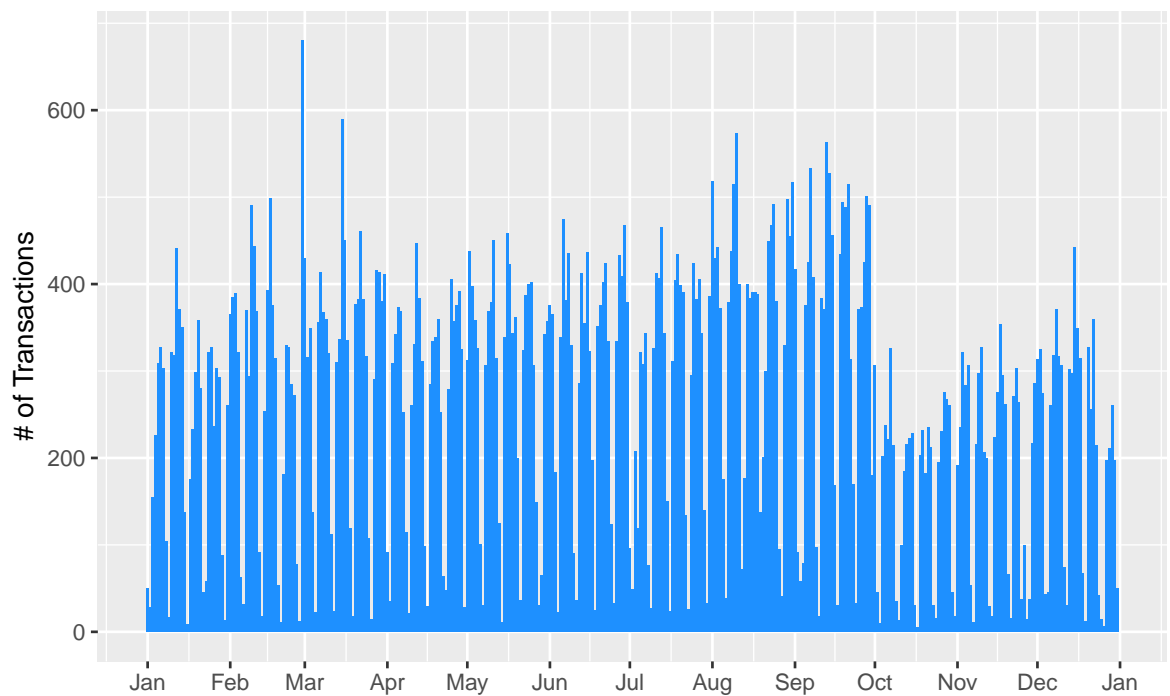
- **Interpretation:** Credit card number of credit card holder.
- **Type:** ID
- **Unique Values:** 1,638
- **Missing Values:** 0 (0.00%)
- **Mean:** NA
- **Standard Deviation:** NA
- **Distribution:** NA

date

- **Interpretation:** Transaction date
- **Type:** Date
- **Unique Values:** 365
- **Missing Values:** 0 (0.00%)
- **Mean:** NA
- **Standard Deviation:** NA
- **Distribution:**

Distribution of date

100.00% of data shown (NA's removed)



merch_num

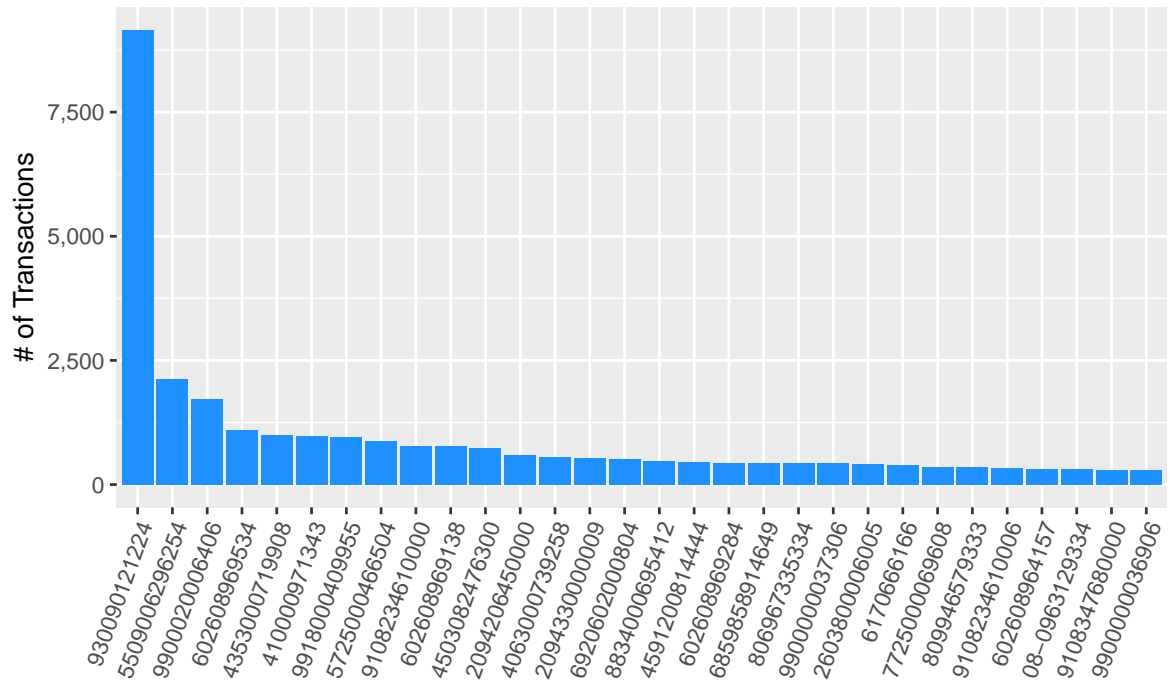
- **Interpretation:** Merchant ID
- **Type:** ID
- **Unique Values:** 13,091
- **Missing Values:** 3,351 (3.51%)
- **Mean:** NA

- **Standard Deviation:** NA

- **Distribution:**

Distribution of merch_num: Top 30

29.42% of data shown (NA's removed)

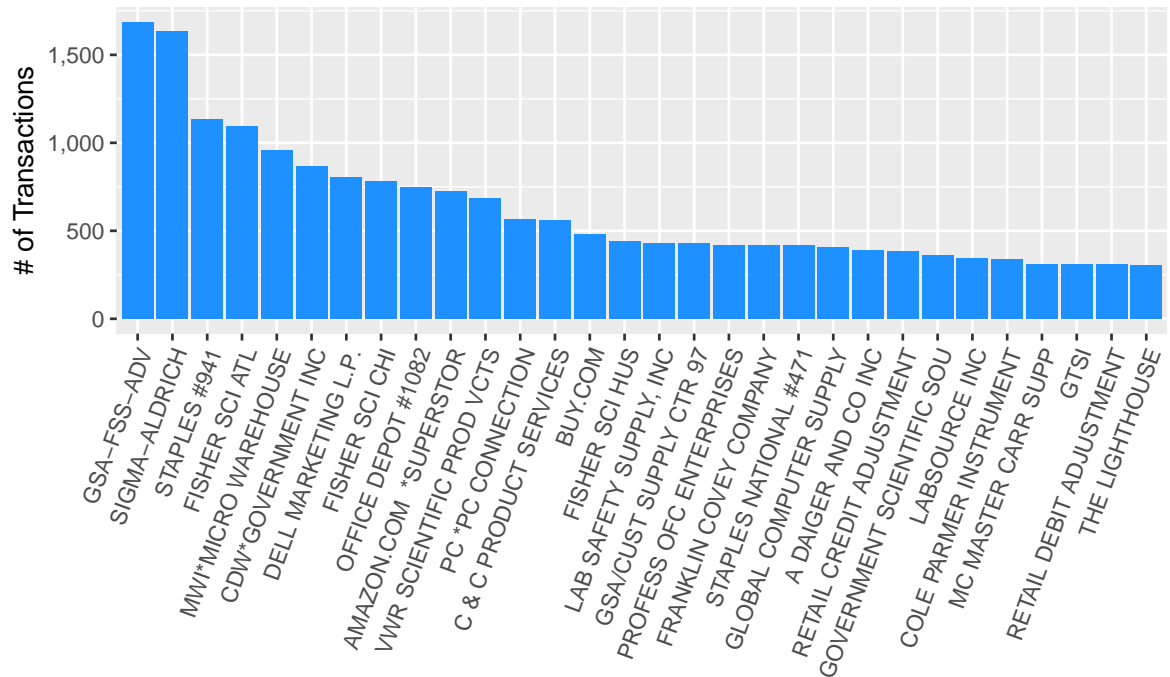


merch_desc

- **Interpretation:** Merchant description
- **Type:** ID
- **Unique Values:** 13,125
- **Missing Values:** 0 (0.00%)
- **Mean:** NA
- **Standard Deviation:** NA
- **Distribution:**

Distribution of merch_desc: Top 30

19.61% of data shown (NA's removed)

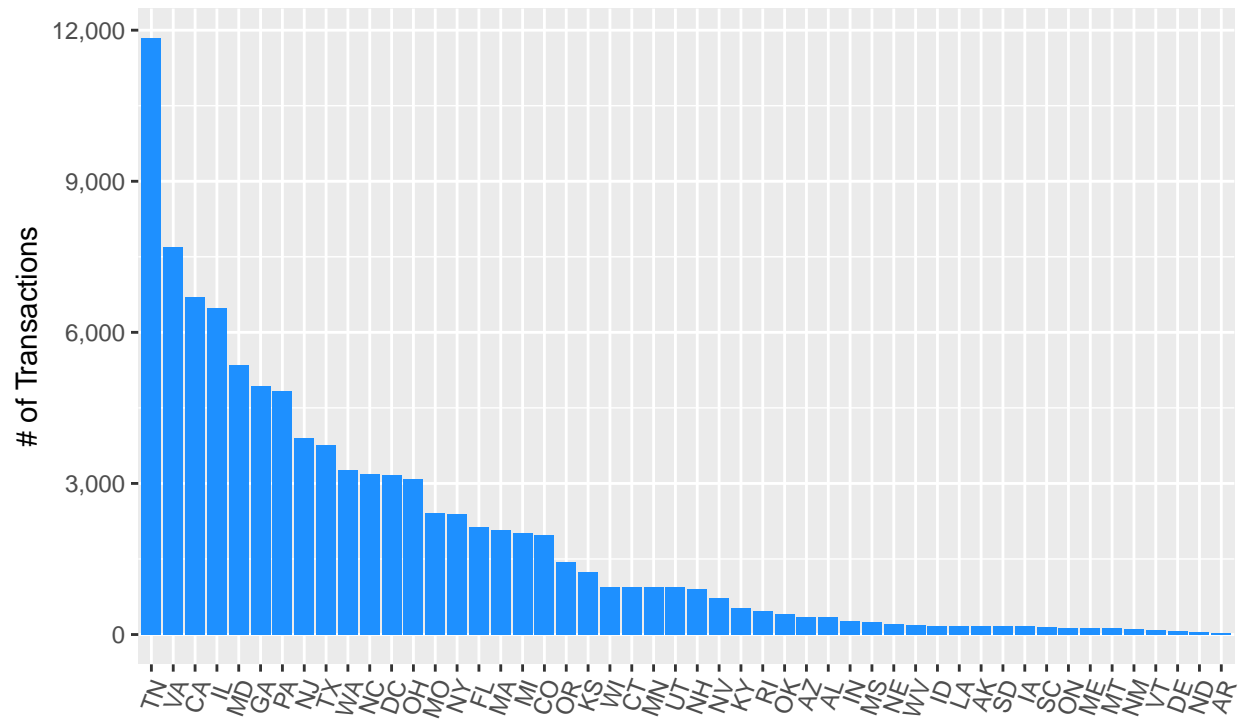


merch_state

- **Interpretation:** Merchant state
- **Type:** Categorical
- **Unique Values:** 228
- **Missing Values:** 1,191 (1.25%)
- **Mean:** NA
- **Standard Deviation:** NA
- **Distribution:**

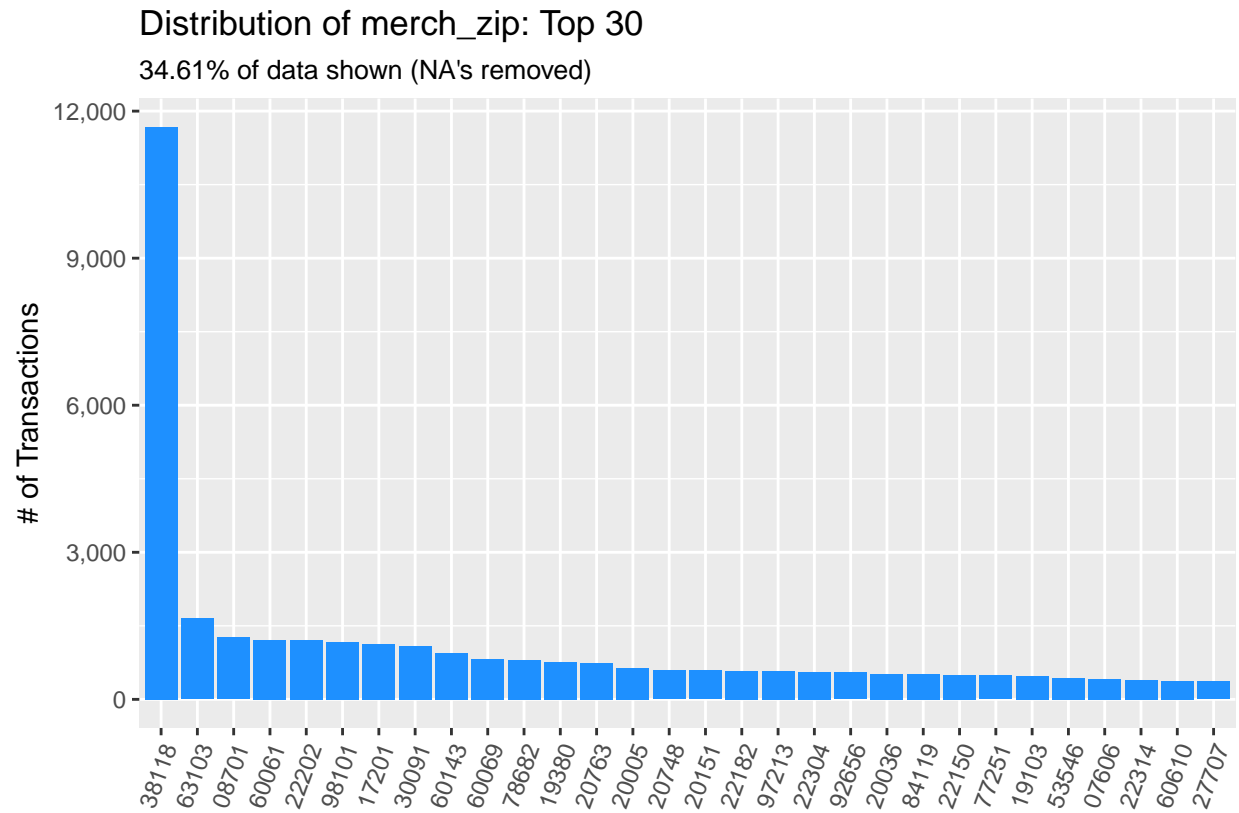
Distribution of merch_state: Top 50

98.47% of data shown (NA's removed)



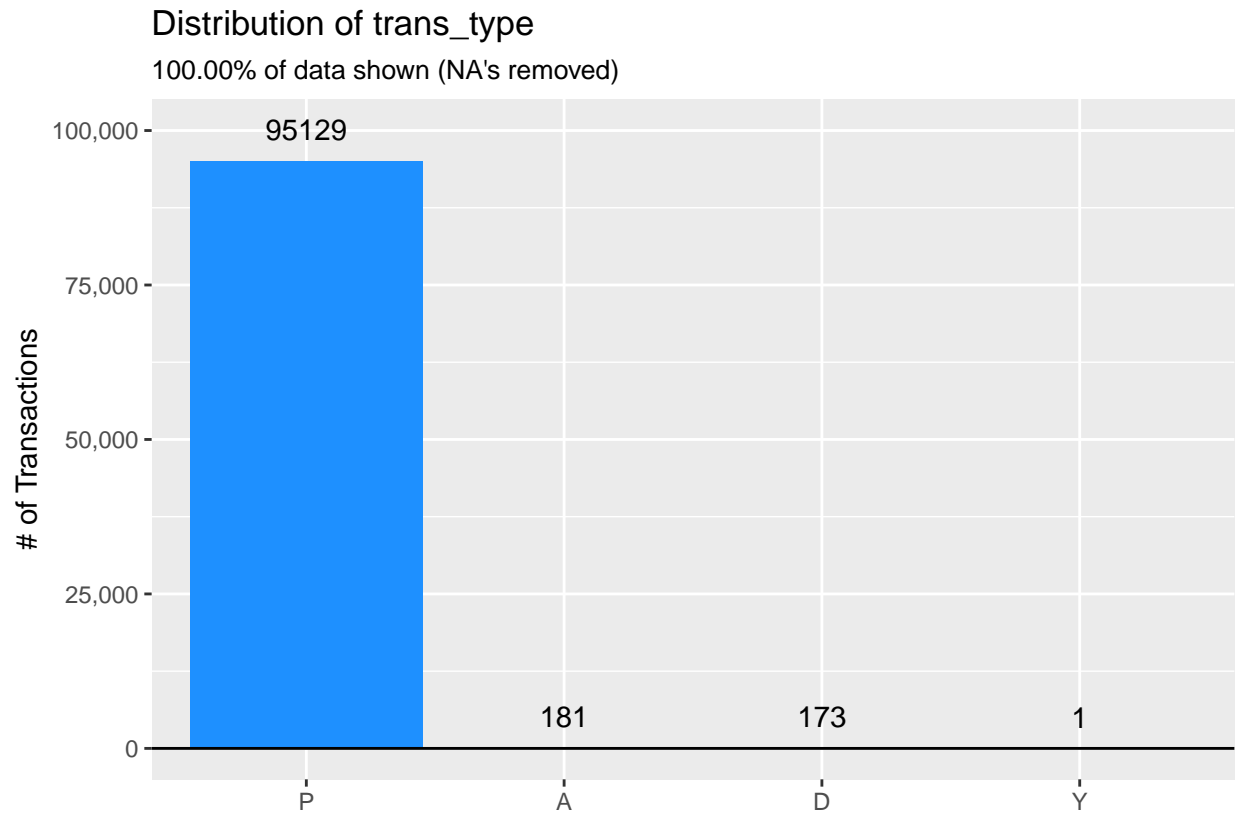
merch_zip

- **Interpretation:** Merchant zipcode
- **Type:** Categorical
- **Unique Values:** 4,568
- **Missing Values:** 4,639 (4.86%)
- **Mean:** NA
- **Standard Deviation:** NA
- **Distribution:**



trans_type

- **Interpretation:** Transaction type
- **Type:** Categorical
- **Unique Values:** 4
- **Missing Values:** 0 (0.00%)
- **Mean:** NA
- **Standard Deviation:** NA
- **Distribution:**

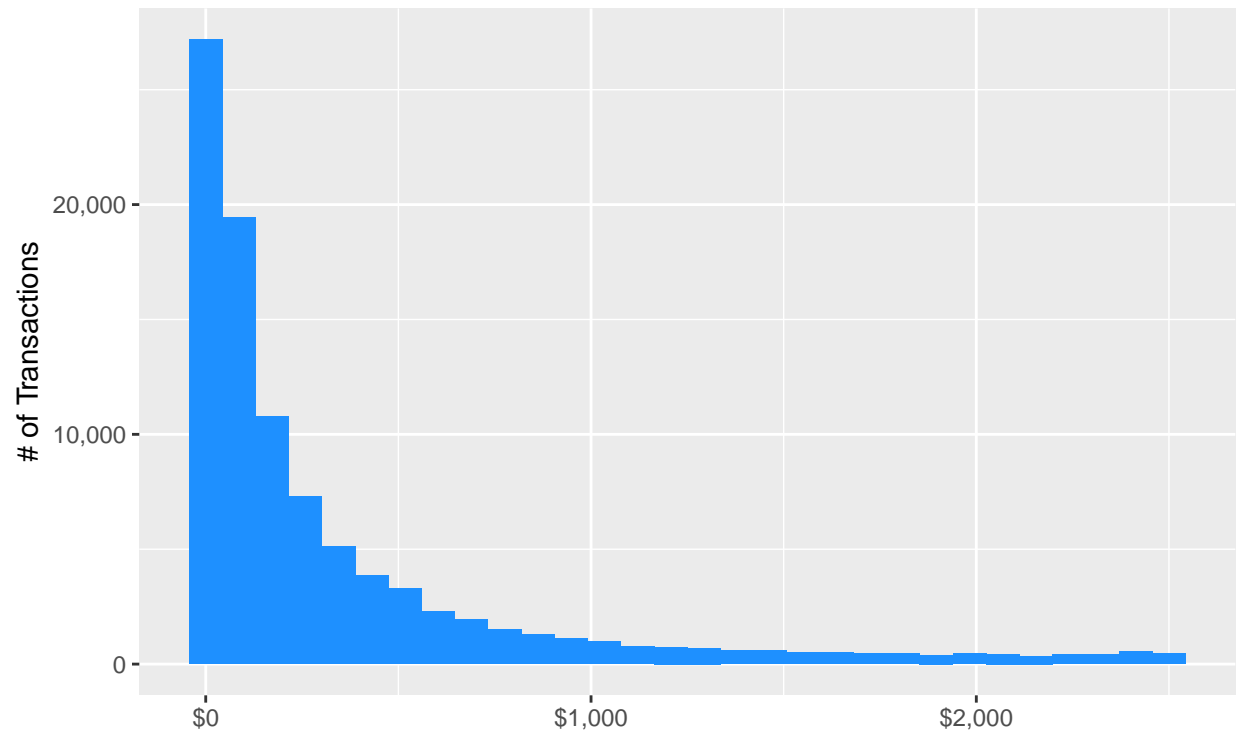


amount

- **Interpretation:** Transaction amount in US dollars (\$)
- **Type:** Numeric
- **Unique Values:** 34,190
- **Missing Values:** 0 (0.00%)
- **Mean:** \$414.38
- **Standard Deviation:** \$10,066.23
- **Min:** \$0.01
- **Max:** \$3,102,045.53
- **Distribution:**

Distribution of amount

99.30% of data shown (NA's removed)



fraud

- **Interpretation:** Indicator whether transaction is fraudulent
- **Type:** Categorical
- **Unique Values:** 2
- **Missing Values:** 0 (0.00%)
- **Mean:** NA
- **Standard Deviation:** NA
- **Distribution:**

