

ANZ Task 1

Hayden Savage

03/08/2021

```
# change extraction column to date and time format

ls = lapply(gsub("T", " ", data$extraction), as.POSIXct)
vec = c()
for (i in ls) {
  vec = append(vec, i)
}

data$extraction = vec

# Average transaction amount
mean(data$amount)

## [1] 187.9336

# Transactions per month
transaction_amoount = data %>%
  mutate(month = format(date, "%m")) %>%
  count(month)

mean(transaction_amoount$n) # Average number of tragnsactions per month

## [1] 4014.333

mean(transaction_amoount$n)/100

## [1] 40.14333

# Age of customers
age_data = data %>%
  group_by(customer_id) %>%
  summarise(age = mean(age))

mean(age_data$age)

## [1] 31.77

sdata = data %>%
  filter(merchant_state != "NA")

state_dat = sdata %>%
  group_by(merchant_state) %>%
  summarise("Proportion of transaction volume (%)") = 100*sum(amount)/sum(sdata$amount))

state_dat
```

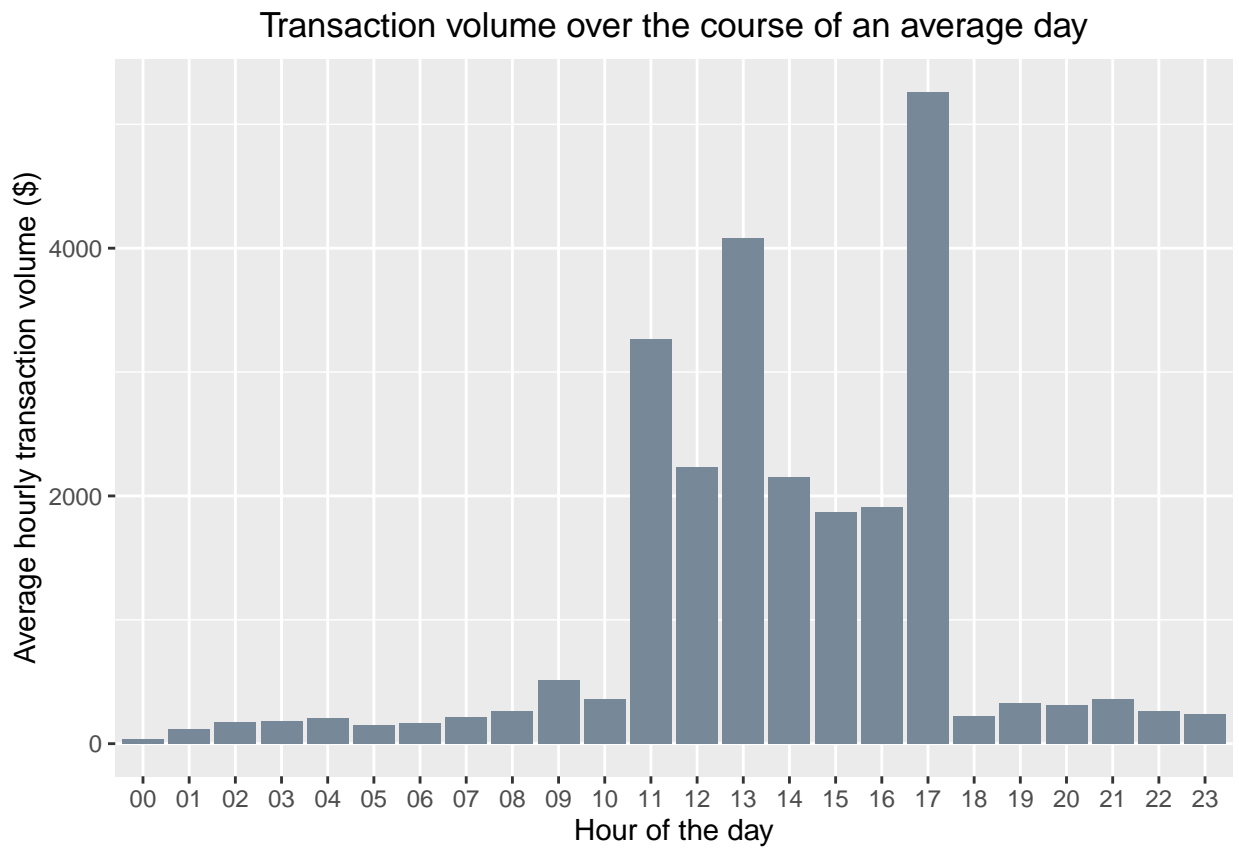
```
## # A tibble: 8 x 2
##   merchant_state `Proportion of transaction volume (%)`
##   <chr>          <dbl>
## 1 ACT           1.57
## 2 NSW           32.9
## 3 NT            2.96
## 4 QLD           17.3
## 5 SA            5.41
## 6 TAS           0.633
## 7 VIC           28.3
## 8 WA            11.0
```

Average transaction amount = \$187.92 Average number of transactions per month per customer = 40 Average customer age = 32 years old

```
num_days = length(unique(data$date))
```

```
hour_data = data %>%
  mutate(hour = format(extraction, "%H")) %>%
  group_by(hour) %>%
  summarise("avg" = sum(amount)/(num_days))
```

```
ggplot(hour_data, aes(hour, avg)) +
  geom_col(fill='lightslategray') +
  labs(x="Hour of the day", y="Average hourly transaction volume ($)", title="Transaction volume over the course of an average day", theme(plot.title = element_text(hjust = 0.5)))
```

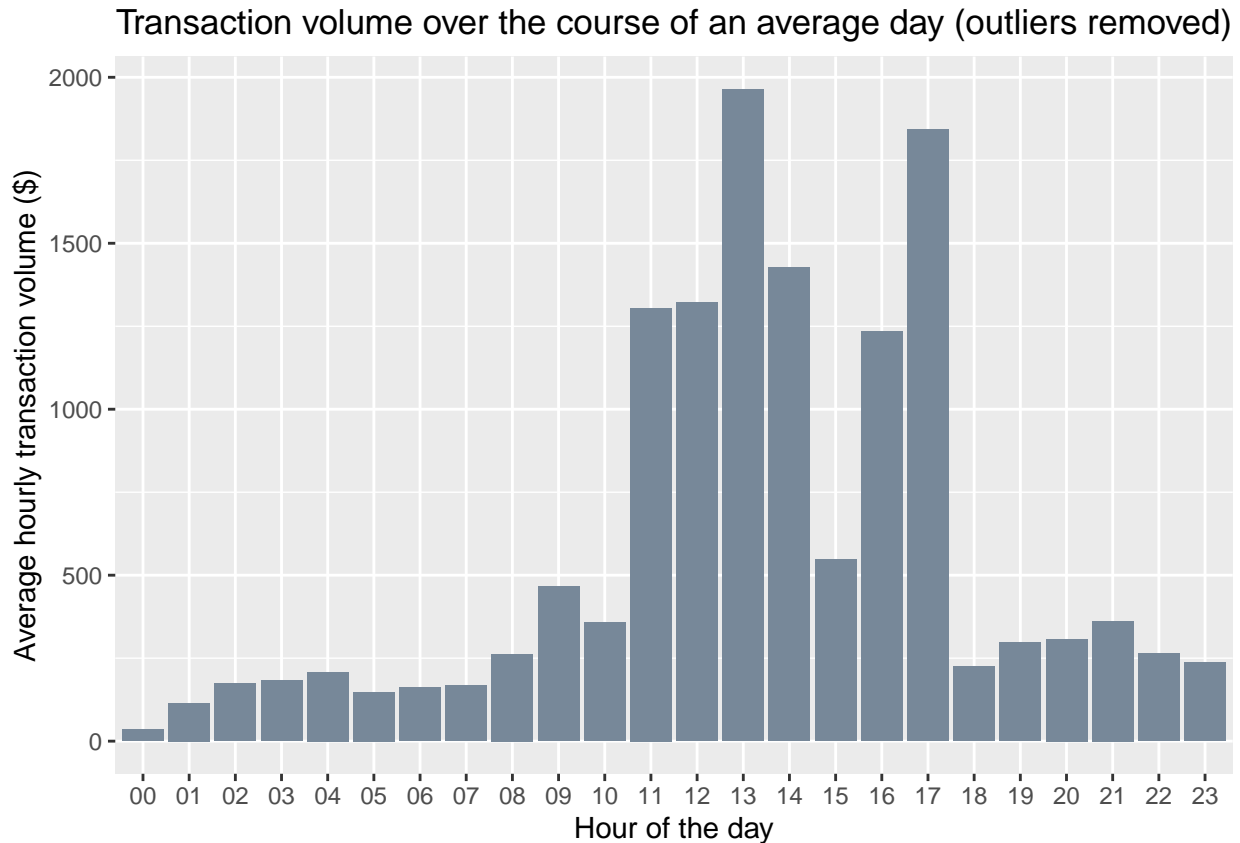


```

hour_data2 = data %>%
  filter(amount < mean(amount) + sd(amount)*3) %>%
  mutate(hour = format(extraction, "%H")) %>%
  group_by(hour) %>%
  summarise("avg" = sum(amount)/(num_days))

ggplot(hour_data2, aes(hour, avg)) +
  geom_col(fill='lightslategray') +
  labs(x="Hour of the day", y="Average hourly transaction volume ($)", title="Transaction volume over the course of an average day (outliers removed)")
  theme(plot.title = element_text(hjust = 0.5))

```



```

day_data = data %>%
  mutate(day = weekdays(date)) %>%
  group_by(day, movement) %>%
  summarise(avg = sum(amount)/(num_days/7))

```

`summarise()` has grouped output by 'day'. You can override using the `.groups` argument.

```

day_data2 = data %>%
  filter(amount < mean(amount) + sd(amount)*3) %>%
  mutate(day = weekdays(date)) %>%
  group_by(day, movement) %>%
  summarise(avg = sum(amount)/(num_days/7))

```

`summarise()` has grouped output by 'day'. You can override using the `.groups` argument.

```

day_levels = c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday")

```

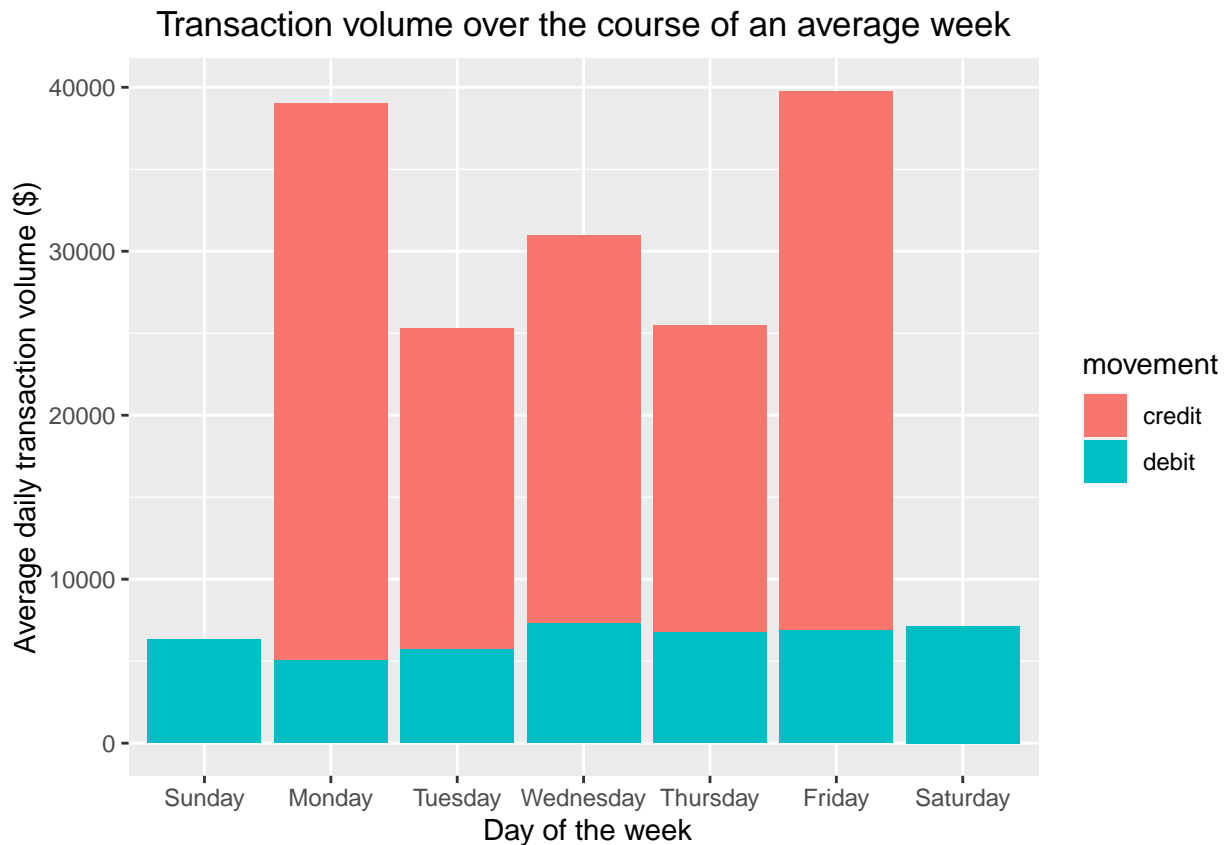
```

day_data = day_data %>%
  mutate(day = factor(day, levels = day_levels))

day_data2 = day_data2 %>%
  mutate(day = factor(day, levels = day_levels))

ggplot(day_data, aes(day, avg, fill=movement)) +
  geom_col() +
  labs(fill="movement", x=" Day of the week", y="Average daily transaction volume ($)", title="Transact.",
  theme(plot.title = element_text(hjust = 0.5))

```

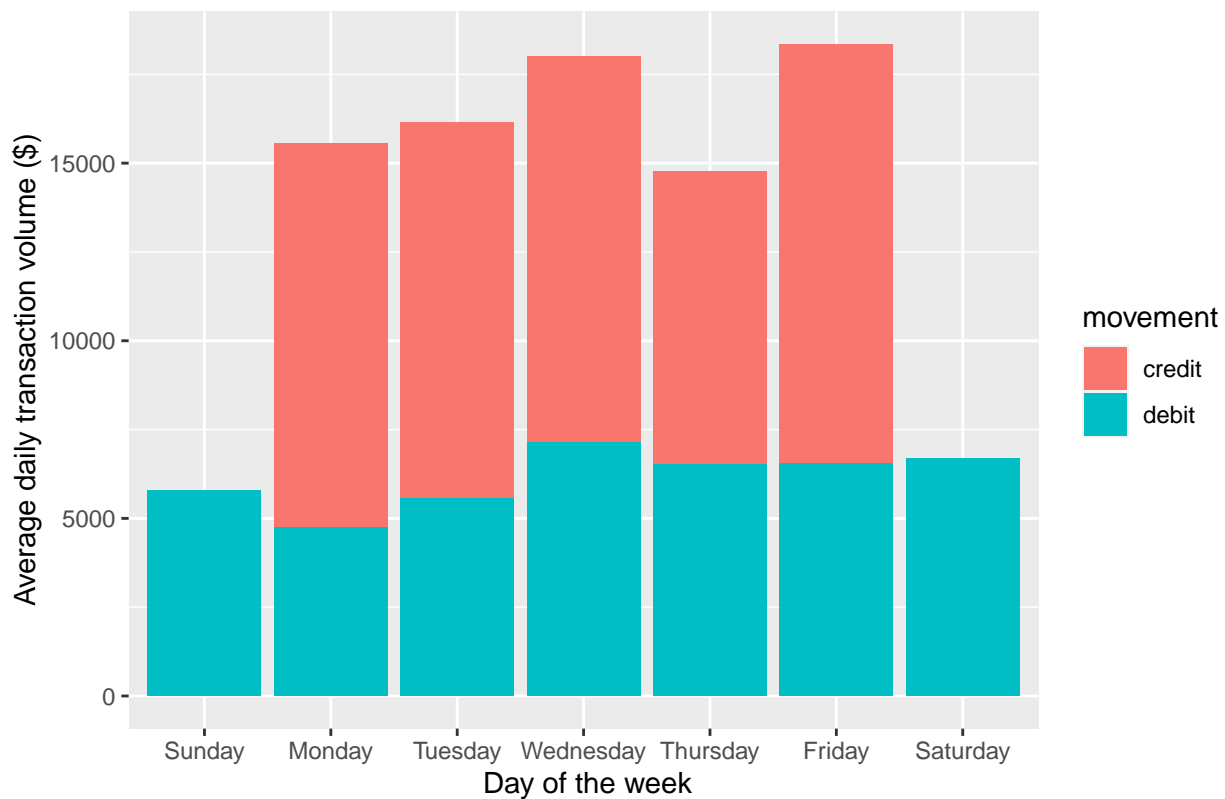


```

ggplot(day_data2, aes(day, avg, fill=movement)) +
  geom_col() +
  labs(fill="movement", x=" Day of the week", y="Average daily transaction volume ($)", title="Transact.",
  theme(plot.title = element_text(hjust = 0.5))

```

Transaction volume over the course of an average week (outliers removed)



```
state_dat = data %>%
  filter(merchant_state != 'NA') %>%
  filter(movement == "debit") %>%
  group_by(merchant_state) %>%
  summarise(avg = mean(amount))
```

```
state_dat
```

```
## # A tibble: 8 x 2
##   merchant_state avg
##   <chr>         <dbl>
## 1 ACT           66.8
## 2 NSW           47.0
## 3 NT            44.7
## 4 QLD           34.4
## 5 SA            40.4
## 6 TAS           28.9
## 7 VIC           41.1
## 8 WA            30.9
```

```
ggplot(state_dat, aes(merchant_state, avg)) +
  geom_col(fill='cyan4') +
  labs(x="State", y="Transaction amount ($)", title="Average transaction amount per state (spending)") +
  theme(plot.title = element_text(hjust = 0.5))
```

