

Kaiyo Warehouse Traffic

Mariel Pacada

5/15/2020

```
setwd("/Users/marielpacada/kaiyo-traffic")
categories <- read.csv("categories.csv")
warehouse <- read.csv("warehouse_flux.csv")

# remove duplicates
warehouse$X <- NULL
warehouse <- unique(warehouse)

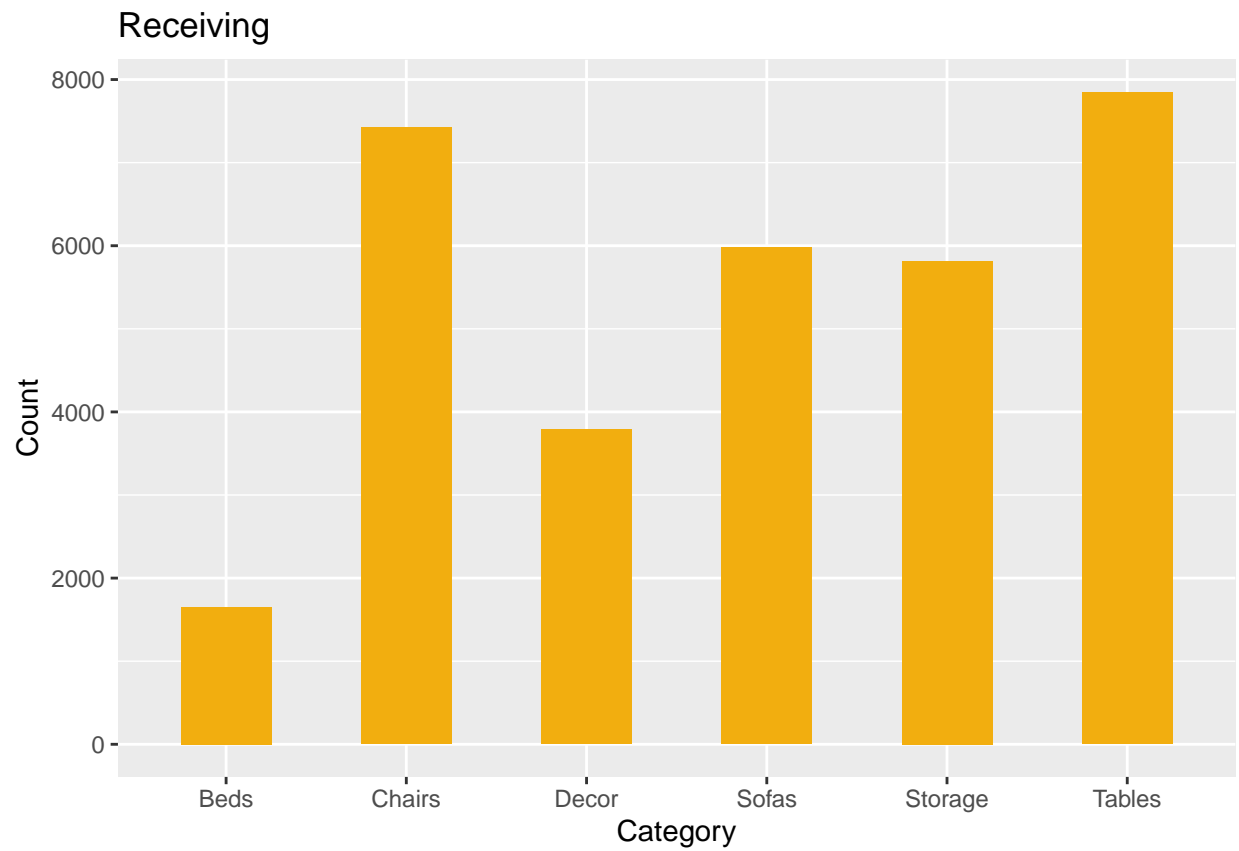
# join warehouse data with item categories
names(categories)[1] <- "category_id"
warehouse <- merge(warehouse, categories, by = "category_id")
warehouse$category_id <- NULL

# order by timestamp
warehouse$create_date <- strptime(warehouse$create_date, format = "%Y-%m-%d %H:%M:%S")
warehouse <- warehouse[order(warehouse$create_date),]

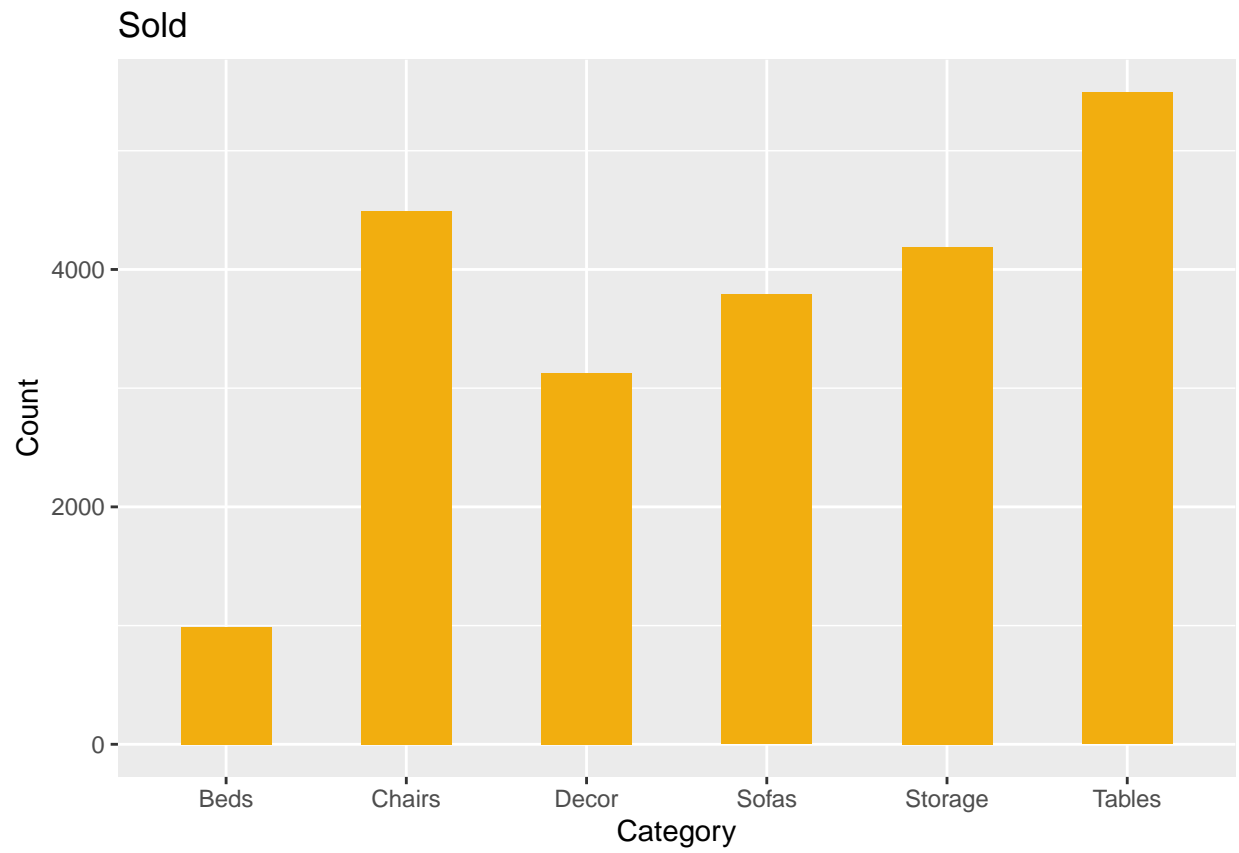
# parse month and year from timestamp + delete timestamp
warehouse$year <- as.factor(warehouse$create_date$year + 1900)
warehouse$month <- warehouse$create_date$mon + 1
warehouse$month <- month.abb[warehouse$month]
warehouse$create_date <- NULL

# function input: factor from status column
# function output: bar chart displaying count by each category
status_count <- function(activity) {
  data <- warehouse %>% filter(status == activity)
  return(ggplot(data, aes(x = category)) + geom_bar(fill = "#F2AE0F", width = 0.5) +
    labs(title = activity) +
    xlab("Category") + ylab("Count"))
}

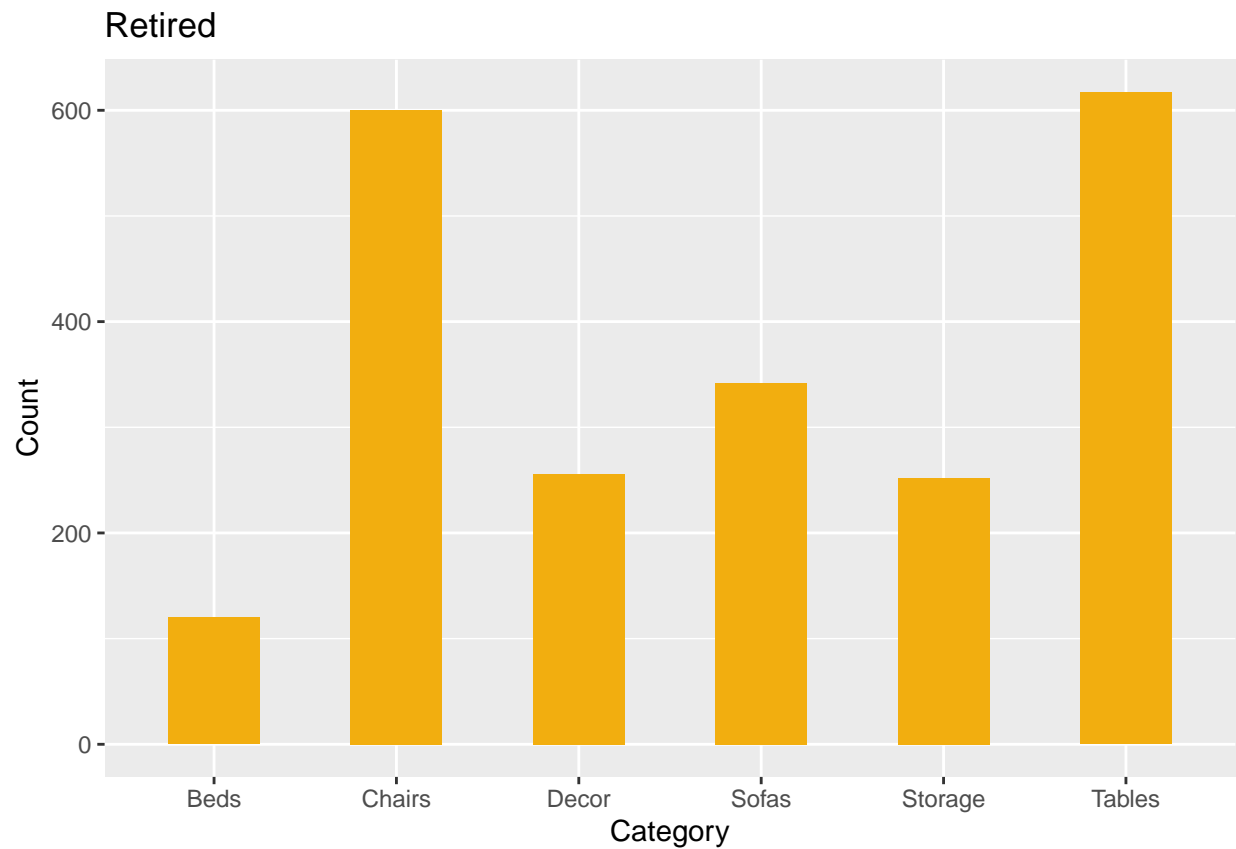
status_count("Receiving")
```



```
status_count("Sold")
```

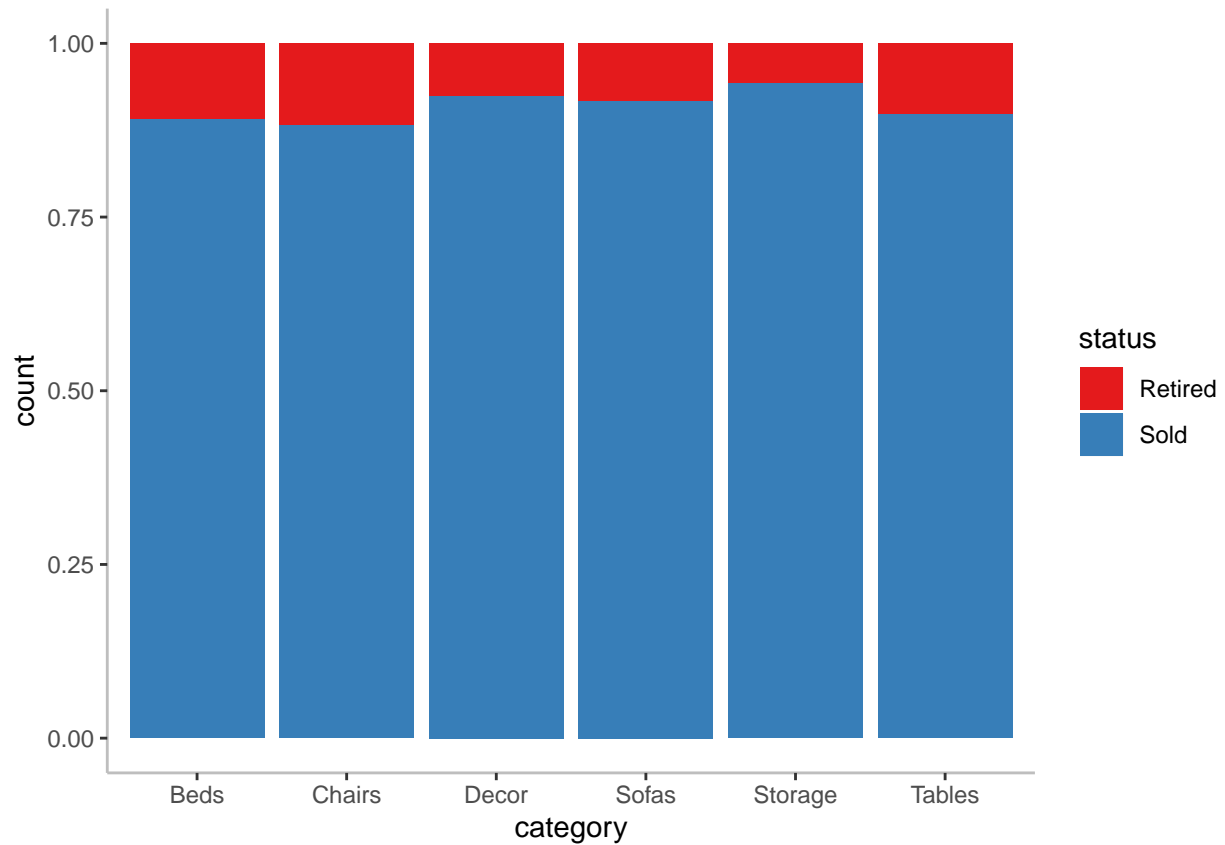


```
status_count("Retired")
```



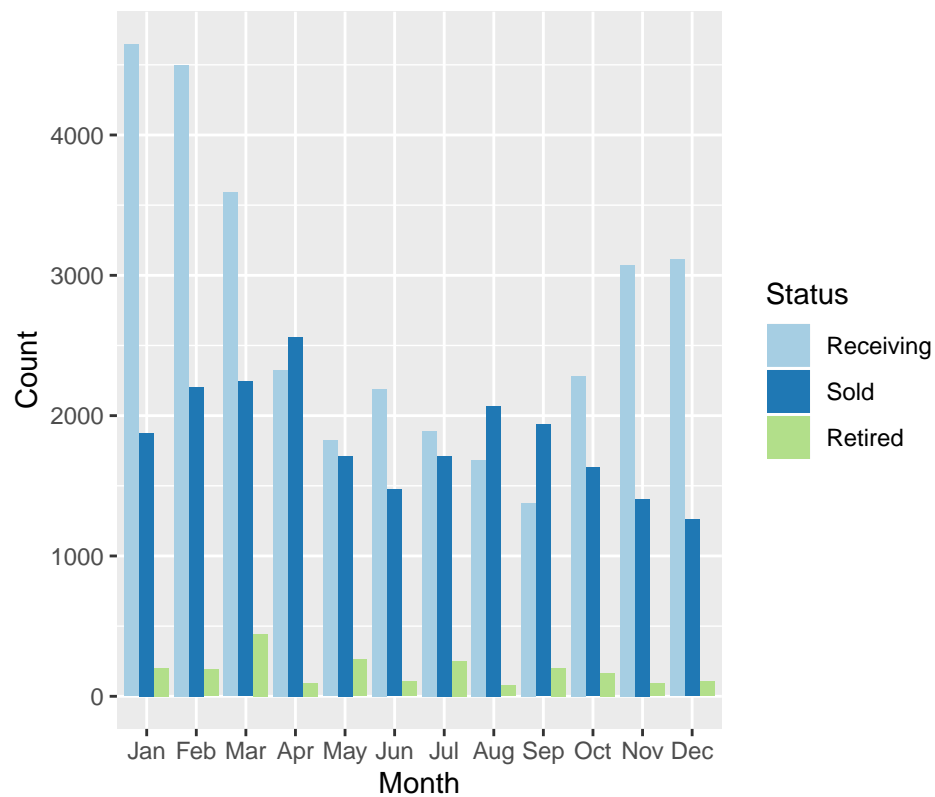
```
# all the items that have left
out_items <- warehouse %>% filter(status != "Receiving")

ggplot(out_items, aes(x = category, fill = status)) + geom_bar(position = "fill") +
  labs() +
  theme(panel.background = element_blank(), axis.line = element_line(colour = "grey")) +
  scale_fill_brewer(palette = "Set1")
```



```
ggplot(warehouse, aes(x = factor(month, levels = month.abb),
                      fill = factor(status, levels = c("Receiving", "Sold", "Retired")))) +
  geom_bar(position = position_dodge()) +
  labs(title = "Monthly Receiving and Selling Activity", fill = "Status") +
  xlab("Month") + ylab("Count") +
  scale_fill_brewer(palette = "Paired")
```

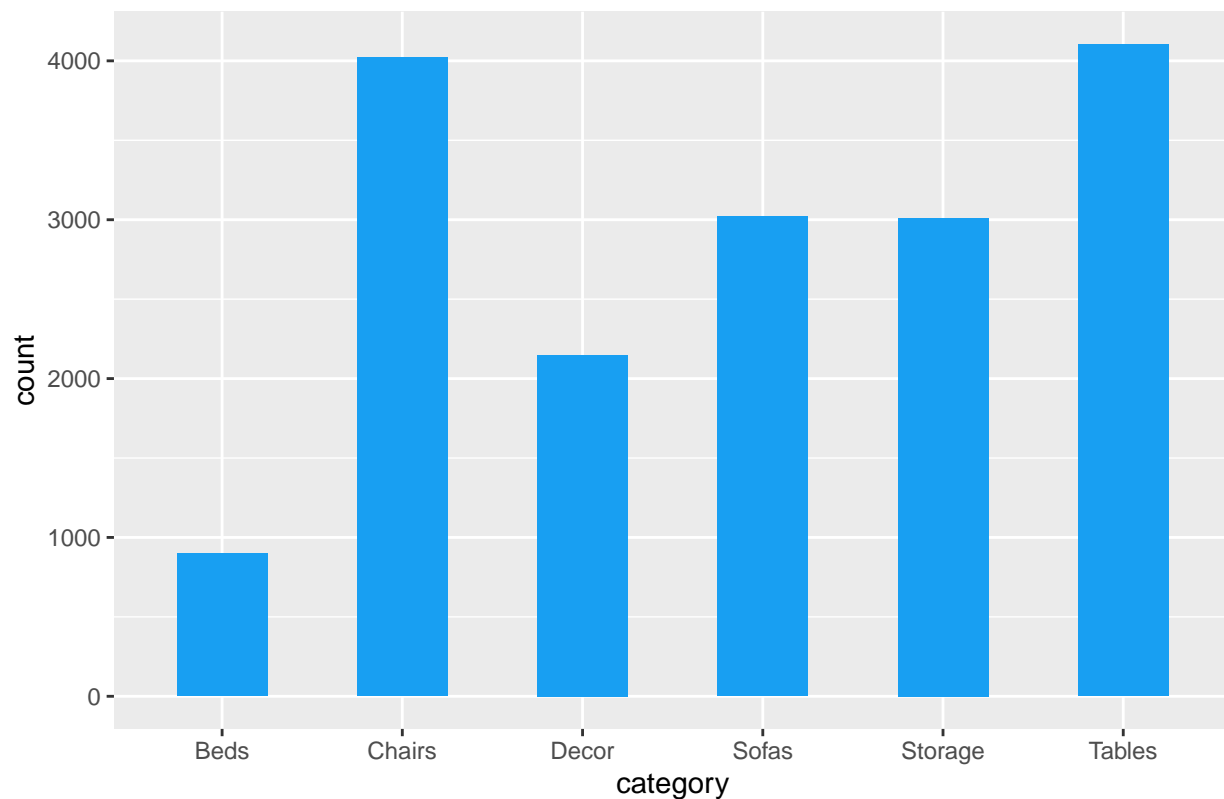
Monthly Receiving and Selling Activity



```
first_tri <- warehouse %>%
  filter(month == "Jan"|month == "Feb"|month == "Mar" & status == "Receiving") %>%
  group_by(category) %>%
  summarize(count = n())

ggplot(first_tri, aes(x = category, y = count)) + geom_bar(fill = "#189ff2", stat = "identity", width = 1) +
  labs(title = "First Trimester Receiving")
```

First Trimester Receiving



```
# some items are classified as two different categories!!!
all_received <- warehouse %>% filter(status == "Receiving") %>% select(status, subitem_id, category)

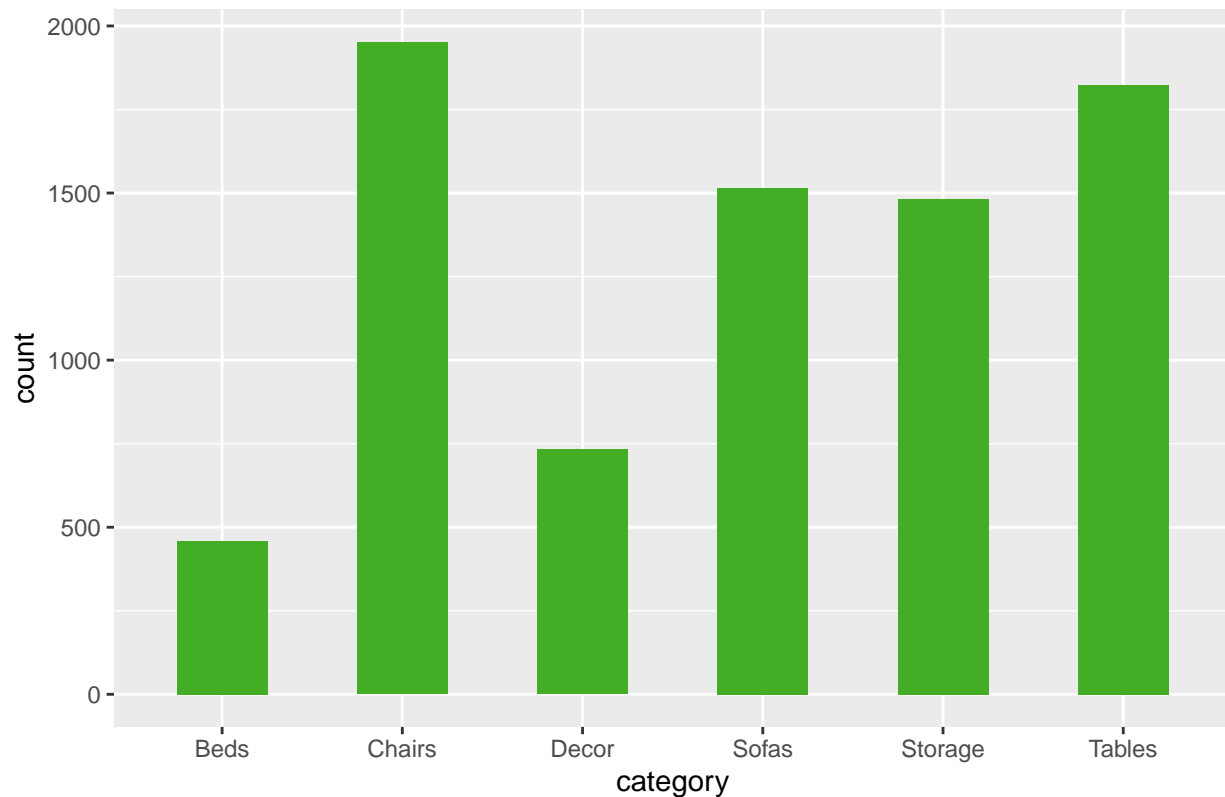
# we check duplication on subitem_id because some items are logged as different categories
once_received <- all_received[!duplicated(all_received$subitem_id),]

# how many times a unique item was received
repeat_received <- all_received %>% group_by(subitem_id) %>% summarize(count = n())
repeat_received <- merge(repeat_received, once_received, by = "subitem_id")

# subsets the items that were received more than once, and then counts how many ITEMS (not how many times)
# re-received) were received more than once for each category
repeat_received <- repeat_received %>%
  filter(count > 1) %>%
  group_by(category) %>%
  summarize(count = n())

ggplot(repeat_received, aes(x = category, y = count)) +
  geom_bar(fill = "#43ad26", stat = "identity", width = 0.5) +
  labs(title = "Which category is most re-received?")
```

Which category is most re-received?



CHAIRS ARE RE-RECEIVED MOST OFTEN -- COMPARE THIS TO SOLD COUNT (IN THE BEGINNING)

```
first_quad <- warehouse %>% filter(month == "Jan"|month == "Feb"|month == "Mar"|month == "Apr")
```

this should be two-thirds of the third_of_year data

```
pre_covid <- first_quad %>% filter(year != "2020")
```

this should be one-third of the third_of_year_data

```
covid <- first_quad %>% filter(year == "2020")
```

```
quad_activity <- function(activity) {
  total <- length(first_quad$status[first_quad$status == activity])
  before <- length(pre_covid$status[pre_covid$status == activity])
  during <- length(covid$status[covid$status == activity])

  before_ratio <- before / total
  during_ratio <- during / total

  return(as.data.frame(t(rbind(c("before", "during"), c(before_ratio, during_ratio)))))
}
```

```
quad_activity("Sold")
```

```
##          V1          V2
## 1 before 0.618173629095823
```



```
## 2 during 0.381826370904177
```

```
quad_activity("Receiving")
```

```
##           V1           V2
## 1 before 0.465834384753304
## 2 during 0.534165615246696
```

```
# baseline 2018 and 2019 data
general_activity <- warehouse %>%
  filter(year != "2020") %>%
  group_by(year, status) %>%
  summarize(count = n())

activity_yearly <- function(activity) {
  data <- general_activity %>% filter(status == activity) %>% select(year, count)
  return(data)
}

# grew by nearly 75%
activity_yearly("Receiving")
```

```
## # A tibble: 2 x 2
## # Groups:   year [2]
##   year count
##   <fct> <int>
## 1 2018   8901
## 2 2019  15552
```

```
# not selling as much :(
activity_yearly("Sold")
```

```
## # A tibble: 2 x 2
## # Groups:   year [2]
##   year count
##   <fct> <int>
## 1 2018   8975
## 2 2019   9712
```

```
# increase by 40%
activity_yearly("Retired")
```

```
## # A tibble: 2 x 2
## # Groups:   year [2]
##   year count
##   <fct> <int>
## 1 2018   783
## 2 2019  1096
```