101_wk4_Adv_Vis

Seung Hyun Sung

11/10/2021

DS4B 101-R: R FOR BUSINESS ANALYSIS —-

ADVANCED BUSINESS PLOTS —-

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## Warning: package 'tibble' was built under R version 4.1.1

## Warning: package 'tidyr' was built under R version 4.1.1

## Warning: package 'forcats' was built under R version 4.1.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 4.1.1

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(tidyquant)
```

```
## Loading required package: PerformanceAnalytics
```

```
## Loading required package: xts
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
##
## Attaching package: 'xts'
```

```
## The following objects are masked from 'package:dplyr':
##
##     first, last
```

```
##
## Attaching package: 'PerformanceAnalytics'
```

```
## The following object is masked from 'package:graphics':
##
##     legend
```

```
## Loading required package: quantmod
```

```
## Loading required package: TTR
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
## == Need to Learn tidyquant? ============================================
## Business Science offers a 1-hour course - Learning Lab #9: Performance Analysis & Portfolio Optimiza
## </> Learn more at: https://university.business-science.io/p/learning-labs-pro </>
```

```
library(formattable)

options(digits = 2, scipen = 99)

bike_orderlines_tbl <- read_rds("~/Desktop/University_business_science/DS4B_101/00_data/bike_sales/data_

glimpse(bike_orderlines_tbl)
```

```
## Rows: 15,644
## Columns: 13
## $ order_date     <dttm> 2011-01-07, 2011-01-07, 2011-01-10, 2011-01-10, 2011-0~
## $ order_id       <dbl> 1, 1, 2, 2, 3, 3, 3, 3, 3, 4, 5, 5, 5, 5, 6, 6, 6, 6, 7~
## $ order_line     <dbl> 1, 2, 1, 2, 1, 2, 3, 4, 5, 1, 1, 2, 3, 4, 1, 2, 3, 4, 1~
## $ quantity       <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1~
## $ price          <dbl> 6070, 5970, 2770, 5970, 10660, 3200, 12790, 5330, 1570,~
## $ total_price    <dbl> 6070, 5970, 2770, 5970, 10660, 3200, 12790, 5330, 1570,~
## $ model          <chr> "Jekyll Carbon 2", "Trigger Carbon 2", "Beast of the Ea~
## $ category_1     <chr> "Mountain", "Mountain", "Mountain", "Mountain", "Road",~
## $ category_2     <chr> "Over Mountain", "Over Mountain", "Trail", "Over Mounta~
## $ frame_material <chr> "Carbon", "Carbon", "Aluminum", "Carbon", "Carbon", "Ca~
## $ bikeshop_name  <chr> "Ithaca Mountain Climbers", "Ithaca Mountain Climbers",~
## $ city           <chr> "Ithaca", "Ithaca", "Kansas City", "Kansas City", "Loui~
## $ state          <chr> "NY", "NY", "KS", "KS", "KY", "KY", "KY", "KY", "KY", "~
```

1.0 Lollipop Chart: Top N Customers

- – Great for showing order

Question: How much purchasing power is in top 5 customers? Goal: Visualize top N customers in terms of Revenue, include cumulative percentage

Data Manipulation

```
n <- 10

top_customers_table <- bike_orderlines_tbl %>%
    select(bikeshop_name, total_price) %>%
    mutate(bikeshop_name = as_factor(bikeshop_name) %>% fct_lump(n = n, w = total_price)) %>%
    group_by(bikeshop_name) %>%
    summarise(revenue = sum(total_price)) %>% ungroup() %>%
    mutate(bikeshop_name = bikeshop_name %>% fct_reorder(revenue)) %>%
    mutate(bikeshop_name = bikeshop_name %>% fct_relevel("Other", after = 0)) %>%
    arrange(desc(bikeshop_name)) %>%

    # Revenue Text
    mutate(revenue_text = scales::dollar(revenue, scale = 1e-6, suffix = "M")) %>%

    # Cumulative Percent
    mutate(cum_pct = cumsum(revenue)/sum(revenue)) %>%
    mutate(cum_pct_text = scales::percent(cum_pct)) %>%

    # Rank
    mutate(rank = row_number()) %>%
    # NA_integer_: NA Values must match the data type. The NA_integer_ adds an NA value to integer data
    mutate(rank = case_when(
        rank == max(rank) ~ NA_integer_,
        TRUE ~ rank
        )) %>%
```

```
    #Label Text
    mutate(label_text = str_glue("Rank: {rank}\nRev: {revenue_text}\nCumPct: {cum_pct_text}"))


top_customers_table
```

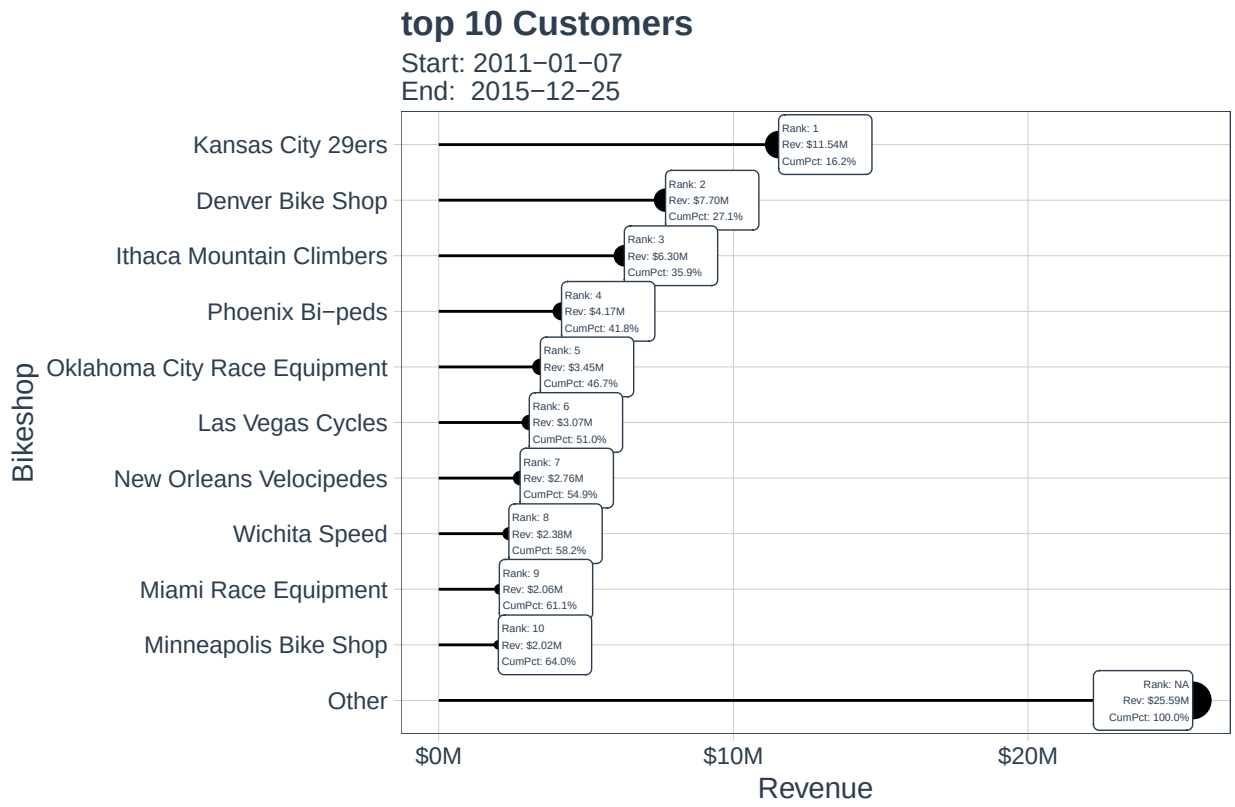
```
## # A tibble: 11 x 7
##    bikeshop_name   revenue revenue_text cum_pct cum_pct_text  rank label_text
##    <fct>             <dbl> <chr>          <dbl> <chr>        <int> <glue>
##  1 Kansas City 29~  1.15e7 $11.54M        0.162 16.2%            1 "Rank: 1\nRe~
##  2 Denver Bike Sh~  7.70e6 $7.70M         0.271 27.1%            2 "Rank: 2\nRe~
##  3 Ithaca Mountai~  6.30e6 $6.30M         0.359 35.9%            3 "Rank: 3\nRe~
##  4 Phoenix Bi-peds  4.17e6 $4.17M         0.418 41.8%            4 "Rank: 4\nRe~
##  5 Oklahoma City ~  3.45e6 $3.45M         0.467 46.7%            5 "Rank: 5\nRe~
##  6 Las Vegas Cycl~  3.07e6 $3.07M         0.510 51.0%            6 "Rank: 6\nRe~
##  7 New Orleans Ve~  2.76e6 $2.76M         0.549 54.9%            7 "Rank: 7\nRe~
##  8 Wichita Speed    2.38e6 $2.38M         0.582 58.2%            8 "Rank: 8\nRe~
##  9 Miami Race Equ~  2.06e6 $2.06M         0.611 61.1%            9 "Rank: 9\nRe~
## 10 Minneapolis Bi~  2.02e6 $2.02M         0.640 64.0%           10 "Rank: 10\nR~
## 11 Other            2.56e7 $25.59M        1     100.0%          NA "Rank: NA\nR~
```

Data Visualization

```
top_customers_table %>%
    # Geometrics
    ggplot(aes(revenue, bikeshop_name)) +
    geom_segment(aes(xend = 0, yend = bikeshop_name), size = 0.5) +
    geom_point(aes(size = revenue)) +

    # Label
    geom_label(aes(label = label_text),
               size = 1.5,
               hjust = "inward",
               colour = palette_light()[1]) +

    # Formatting
    scale_x_continuous(labels = scales::dollar_format(scale = 1e-6, suffix = "M")) +
    labs(
        title = str_glue("top {n} Customers"),
        subtitle = str_glue("Start: {min(bike_orderlines_tbl$order_date)}
                             End:  {max(bike_orderlines_tbl$order_date)}"),
        x = "Revenue",
        y = "Bikeshop",
        caption = str_glue("Top 6 customers contribute 51% of purchasing power.")
    ) +
    theme_tq() +
    theme(
        legend.position = "none",
        plot.title = element_text(face = "bold"),
        plot.caption = element_text(face = "bold.italic")
    )
```



Top 6 customers contribute 51% of purchasing power.

2.0 Heatmaps

- – Great for showing details in 3 dimensions

Question: Do specific customers have a purchasing prefernce?

Goal: Visualize heatmap of proportion of sales by Secondary Product Category

Data Manipulation

```
pct_sales_by_customer <- bike_orderlines_tbl %>%

    select(bikeshop_name, category_1, category_2, quantity) %>%

    group_by(bikeshop_name, category_1, category_2) %>%
    summarise(total_qty = sum(quantity)) %>% ungroup() %>%

    group_by(bikeshop_name) %>%
    mutate(pct = total_qty/sum(total_qty)) %>% ungroup() %>%

    mutate(bikeshop_name = as_factor(bikeshop_name) %>% fct_rev()) %>%
    mutate(bikeshop_name_num = as.numeric(bikeshop_name))
```

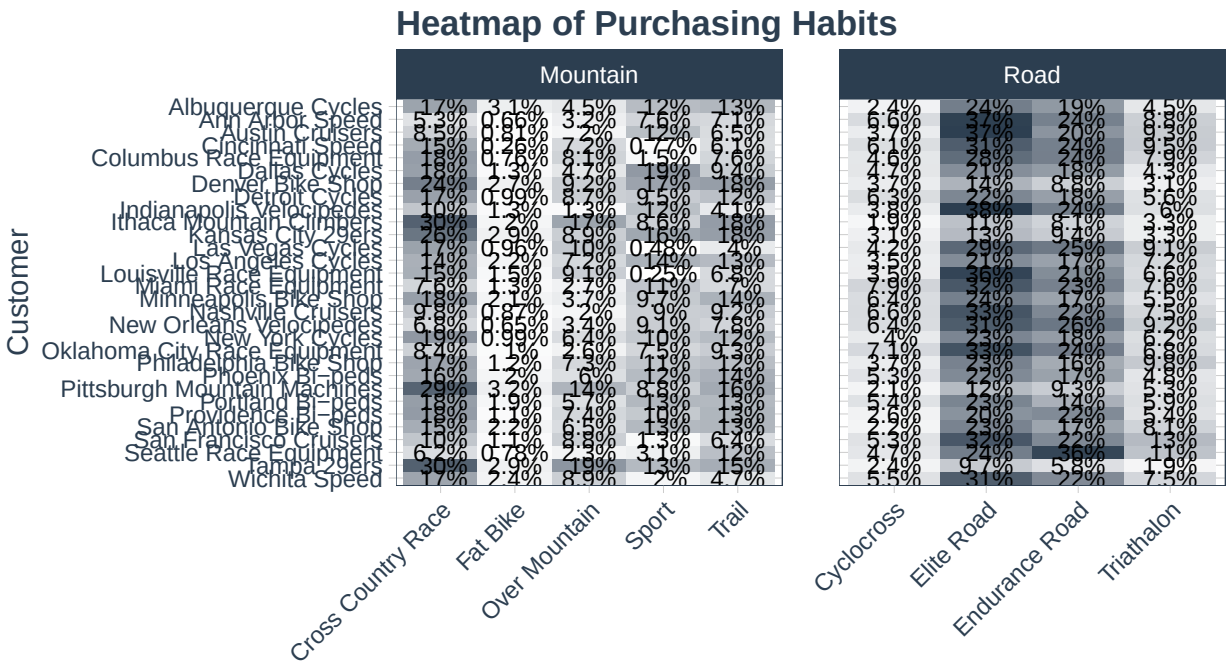
```
## `summarise()` has grouped output by 'bikeshop_name', 'category_1'. You can override using the `.grou
```

Data Visualization

```
pct_sales_by_customer %>%
    ggplot(aes(category_2, bikeshop_name)) +
    geom_tile(aes(fill = pct)) +
    geom_text(aes(label = scales::percent(formattable(pct, digits =2))), size = 3) +

    facet_wrap(~category_1, scales = "free_x") +
    # formatting
    scale_fill_gradient(low = "white", high = palette_light()[1]) +
    labs(
        title = "Heatmap of Purchasing Habits",
        x = "Bike Type (Category 2)",
        y = "Customer",
        caption = str_glue(
        "Customers that prefer Road:
        Ann Arbor speed, Austin Cruisers, & Indianapolis Velocipedes

        Customers that prefer Mountain:
        Ithaca Mountain Climbers, Pittsburgh Mountain Machines, & Tampa 29ers")
    ) +
    theme_tq() +
    theme(
        axis.text.x = element_text(angle = 45, hjust = 1),
        legend.position = "none",
        plot.title = element_text(face = "bold"),
        plot.caption = element_text(face = "bold.italic")
    )
```



Customers that prefer Road:
Ann Arbor speed, Austin Cruisers, & Indianapolis Velocipedes

Customers that prefer Mountain:
Ithaca Mountain Climbers, Pittsburgh Mountain Machines, & Tampa 29ers