**Template for Project-1: SQL & RStudio**

Group membership (MAX 4 students):

* RONGJIA JING

Please note: For each of the queries, you should paste (a) the actual SQL code you used to create the query and (b) the results of your query.

**⚠** NOTE: some of queries generate hundreds of lines; in those cases, copy and paste **just the first 15 rows** of the output table unless otherwise instructed.

*In the event that you have a query that will not run successfully, paste the SQL code as indicated and then copy and paste the last error message that you obtain in place of the table.*

1. (5 points) Now, on your own, modify the query, using the ORDER BY command to list the same results sorted by number of SKUs in each department, beginning with the smallest number of SKUs. Just copy & paste the 15 smallest departments.

(NOTE: This query will generate 333 rows of data plus one title row; just copy/paste the first 15 rows)

SELECT dept, COUNT(sku)

FROM SKUINFO

GROUP BY dept

ORDER BY COUNT(sku)

|  |  |  |
| --- | --- | --- |
| **No** | DEPT | Count(SKU) |
| **1** | 4400 | 269 |
| **2** | 8002 | 275 |
| **3** | 6400 | 286 |
| **4** | 8000 | 324 |
| **5** | 7200 | 330 |
| **6** | 3100 | 1019 |
| **7** | 1100 | 1121 |
| **8** | 5100 | 1449 |
| **9** | 5506 | 1451 |
| **10** | 5207 | 1474 |
| **11** | 3107 | 1818 |
| **12** | 9000 | 2107 |
| **13** | 800 | 2145 |
| **14** | 2200 | 3731 |
| **15** | 6107 | 3929 |

窗体底端

1. (20 pts) Write a query to obtain total number of stores in each state. The table should have two columns: STATE, and Count(STORE). In the Answer Set tab, find the disk icon to save the query table as a Comma Separated Values (csv) format file. Then open that CSV file in RStudio, and develop a geographic map of US states based on number of stores in each state and paste the map below.

SELECT state, COUNT(store) "Number of Stores"

FROM Store\_msa

GROUP BY state

<http://127.0.0.1:16191/custom/googleVis/GeoMapIDf056c570516.html>

图片包含 文字, 地图

描述已自动生成

|  |  |
| --- | --- |
| **STATE** | **Number of Stores** |
| **MT** | 3 |
| **CA** | 3 |
| **OH** | 21 |
| **ID** | 3 |
| **IL** | 3 |
| **NM** | 6 |
| **AR** | 9 |
| **TX** | 57 |
| **KS** | 8 |
| **TN** | 15 |
| **VA** | 10 |
| **SC** | 8 |
| **MS** | 6 |
| **NE** | 4 |
| **UT** | 6 |

3.( 20 pts) How many transactions took place in each of the Dallas, Texas (CITY=’DALLAS’) stores on December 24, 2004 (SALEDATE = ‘2004-12-24’)? What was the dollar value of sales in each store on that date? Your result should be a 3-column table showing the STORE number, transaction count, and SUM of AMT.

Again, just paste the first 10 rows.

SELECT count(distinct trnsact.trannum), trnsact.store, sum(trnsact.amt)

FROM trnsact

INNER JOIN strinfo ON strinfo.store = trnsact.store

WHERE strinfo.city='Dallas' AND trnsact.SALEDATE = '2004-12-24'

GROUP BY trnsact.store

|  |  |  |
| --- | --- | --- |
| **Count(Distinct(TRANNUM))** | **STORE** | **Sum(AMT)** |
| **206** | 1707 | 81199.69 |
| **165** | 707 | 66765.00 |
| **194** | 1607 | 108711.89 |

Next, obtain data on total sales and number of stores in each state. Transfer data in a CSV format to RStudio and develop a geographic map based on both total sales and number of stores in each state using a single map. Paste geographic map below.

SELECT state, SUM(trnsact.amt) "total sales", COUNT(trnsact.store) "Number of Stores"

FROM trnsact

INNER JOIN store\_msa ON trnsact.store = store\_msa.store

GROUP BY state

<http://127.0.0.1:24890/custom/googleVis/GeoMapID2e075f69109.html>

图片包含 文字, 地图

描述已自动生成

1. (10 pts) Let’s look at the transactions more closely: Store #707 in Dallas had 165 transactions that day. Some customers bought 1 item, some bought many. For each transaction (TRANNUM), how many different SKUs were purchased, and what was the dollar sales total for each transaction? List the transactions starting with the largest dollar value. (Hint: your result should also have 165 rows!) Just list the 25 largest sales.

SELECT TOP 10 trnsact.trannum"No.transaction", count(trnsact.sku)"No.sku", sum(trnsact.amt) "total sales"

FROM trnsact

INNER JOIN strinfo ON strinfo.store = trnsact.store

WHERE strinfo.city='Dallas' AND trnsact.SALEDATE = '2004-12-24' AND trnsact.store = '707'

GROUP BY trnsact.trannum

ORDER BY sum(trnsact.amt) desc

|  |  |  |
| --- | --- | --- |
| **No.transaction** | **No.sku** | **total sales** |
| **800** | 43 | 1566.70 |
| **3700** | 29 | 1328.17 |
| **1100** | 40 | 1291.92 |
| **3300** | 33 | 1216.17 |
| **4400** | 31 | 1205.69 |
| **1700** | 36 | 1099.95 |
| **5600** | 37 | 1089.64 |
| **3100** | 29 | 1085.60 |
| **900** | 34 | 1076.95 |
| **700** | 37 | 1067.04 |

1. (10 pts). Finally, the executives responsible for footwear & accessories want to compare the sales performance of two popular brands: Nine West and Nina. Devise a query to compare (state by state) total revenue from the two brands in Dillard’s stores. These totals should be aggregated and broken out by state. Your result should be a table listing all states with Dillard’s stores, and total Nina sales amounts for all 29 states, then Nine West sales amounts for all 29 states. In other words, the top 29 rows should show Nina sales, and the next 29 Nine West sales. In the database, the brands are identified as NINA FOO and NINE WES. As with the other queries, be sure to include Purchases only (not returns)

The query should produce four columns: State, Brand, Total Revenue and Number of Stores. Please list the states in alphabetical order. Your query should generate 58 rows. Please show all of them.

SELECT state, k.brand, sum(amt) "Total rev.", count(t.store) "No. Stores"

FROM skuinfo k

INNER JOIN trnsact t ON t.sku = k.sku

INNER JOIN strinfo x ON x.store = t.store

WHERE k.brand LIKE '%NINA FOO%' OR k.brand LIKE '%NINE WES%'

GROUP BY state, k.brand

ORDER BY k.brand, state

|  |  |  |  |
| --- | --- | --- | --- |
| **STATE** | **BRAND** | **Total rev.** | **No. Stores** |
| **AL** | NINA FOO | 900805.47 | 22418 |
| **AR** | NINA FOO | 554466.75 | 11797 |
| **AZ** | NINA FOO | 1008823.41 | 24677 |
| **CA** | NINA FOO | 266544.66 | 5522 |
| **CO** | NINA FOO | 691895.17 | 15496 |
| **FL** | NINA FOO | 4540766.52 | 101960 |
| **GA** | NINA FOO | 963049.49 | 23964 |
| **IA** | NINA FOO | 228566.47 | 5845 |
| **ID** | NINA FOO | 106116.28 | 2598 |
| **IL** | NINA FOO | 126744.99 | 3118 |
| **IN** | NINA FOO | 76512.06 | 2483 |
| **KS** | NINA FOO | 484383.26 | 13343 |
| **KY** | NINA FOO | 671130.26 | 17096 |
| **LA** | NINA FOO | 1741248.31 | 39326 |
| **MO** | NINA FOO | 876351.95 | 23506 |
| **MS** | NINA FOO | 353988.78 | 8416 |
| **MT** | NINA FOO | 88461.22 | 2266 |
| **NC** | NINA FOO | 619643.94 | 13954 |
| **NE** | NINA FOO | 254364.86 | 6815 |
| **NM** | NINA FOO | 292477.41 | 7383 |
| **NV** | NINA FOO | 380801.12 | 8099 |
| **OH** | NINA FOO | 1566467.37 | 37568 |
| **OK** | NINA FOO | 601437.99 | 13769 |
| **SC** | NINA FOO | 492307.89 | 10365 |
| **TN** | NINA FOO | 802022.01 | 18146 |
| **TX** | NINA FOO | 6342780.53 | 158064 |
| **UT** | NINA FOO | 200383.55 | 5487 |
| **VA** | NINA FOO | 437391.35 | 9291 |
| **WY** | NINA FOO | 29310.06 | 712 |
| **AL** | NINE WES | 842227.66 | 23669 |
| **AR** | NINE WES | 591061.78 | 14021 |
| **AZ** | NINE WES | 1585354.53 | 47055 |
| **CA** | NINE WES | 366403.24 | 9759 |
| **CO** | NINE WES | 719237.03 | 22829 |
| **FL** | NINE WES | 3101979.38 | 93513 |
| **GA** | NINE WES | 889267.77 | 25930 |
| **IA** | NINE WES | 304675.18 | 10581 |
| **ID** | NINE WES | 159126.62 | 5181 |
| **IL** | NINE WES | 160074.29 | 5489 |
| **IN** | NINE WES | 125559.20 | 4146 |
| **KS** | NINE WES | 589865.85 | 18110 |
| **KY** | NINE WES | 709878.71 | 22851 |
| **LA** | NINE WES | 1469414.86 | 40680 |
| **MO** | NINE WES | 942689.25 | 31824 |
| **MS** | NINE WES | 343515.67 | 9551 |
| **MT** | NINE WES | 95812.46 | 3339 |
| **NC** | NINE WES | 804486.01 | 23032 |
| **NE** | NINE WES | 369365.86 | 11193 |
| **NM** | NINE WES | 458567.54 | 13820 |
| **NV** | NINE WES | 633862.32 | 17823 |
| **OH** | NINE WES | 1408939.83 | 45931 |
| **OK** | NINE WES | 749882.02 | 18183 |
| **SC** | NINE WES | 547783.01 | 14487 |
| **TN** | NINE WES | 940140.35 | 24745 |
| **TX** | NINE WES | 3988336.53 | 119191 |
| **UT** | NINE WES | 292551.49 | 9138 |
| **VA** | NINE WES | 475755.22 | 14033 |
| **WY** | NINE WES | 32053.05 | 1098 |

1. (30 pts). In the CSV file (from step 5) you have 2 rows for each state, with sales amounts corresponding to two different brands. Rearrange your CSV worksheet to have three columns: State, Nina, and NineWest.  
     
   Prepare displays as mentioned along with R codes used:
   1. A map each for NINA FOO and NINE WES showing states with Dillard’s stores, using colors and/or sizes to represent both the number of stores in the state and Total Revenue.

NINAFOO$STATE <-abbr2state(NINAFOO$STATE)

G3<-gvisGeoMap(NINAFOO,locationvar = 'STATE',

numvar = 'No. Stores',

hovervar = 'Total rev.',

options = list(region='US',dataMode="regions"))

plot(G3)

NINEWES$STATE <-abbr2state(NINEWES$STATE)

G4<-gvisGeoMap(NINEWES,locationvar = 'STATE',

numvar = 'No. Stores',

hovervar = 'Total rev.',

options = list(region='US',dataMode="regions"))

plot(G4)

图片包含 文字, 地图

描述已自动生成

图片包含 文字, 地图

描述已自动生成

* 1. Four additional display of your choosing that clearly summarizes and presents appropriate information.

par(mfrow=c(2,1))

boxplot(NINAFOO$`No. Stores`~NINAFOO$STATE,

NINAFOO,

xlab = "STATE", ylab = "Number of stores(NINAFOO)",ylim = c(0,150000),

las =2)

boxplot(NINEWES$`No. Stores`~NINEWES$STATE,

NINEWES,

xlab = "STATE", ylab = "Number of stores(NINEWES)",ylim = c(0,150000),

las =2)

par(mfrow=c(2,1))

boxplot(NINAFOO$`Total rev.`~NINAFOO$STATE,

NINAFOO,

xlab = "STATE", ylab = "Total Revenue(NINAFOO)",ylim = c(0,1000000),

las =2)

boxplot(NINEWES$`Total rev.`~NINEWES$STATE,

NINEWES,

xlab = "STATE", ylab = "Total Revenue(NINEWES)",ylim = c(0,1000000),

las =2)

RevperStore <-NINAFOO$`Total rev.`/NINAFOO$`No. Stores`

RevperStore2 <-NINEWES$`Total rev.`/NINEWES$`No. Stores`

par(mfrow=c(2,1))

boxplot(RevperStore~NINAFOO$STATE,

xlab = "STATE", ylab = "Revenue per Store(NINAFOO)",ylim = c(0,60),

las =2)

boxplot(RevperStore2 ~NINEWES$STATE,

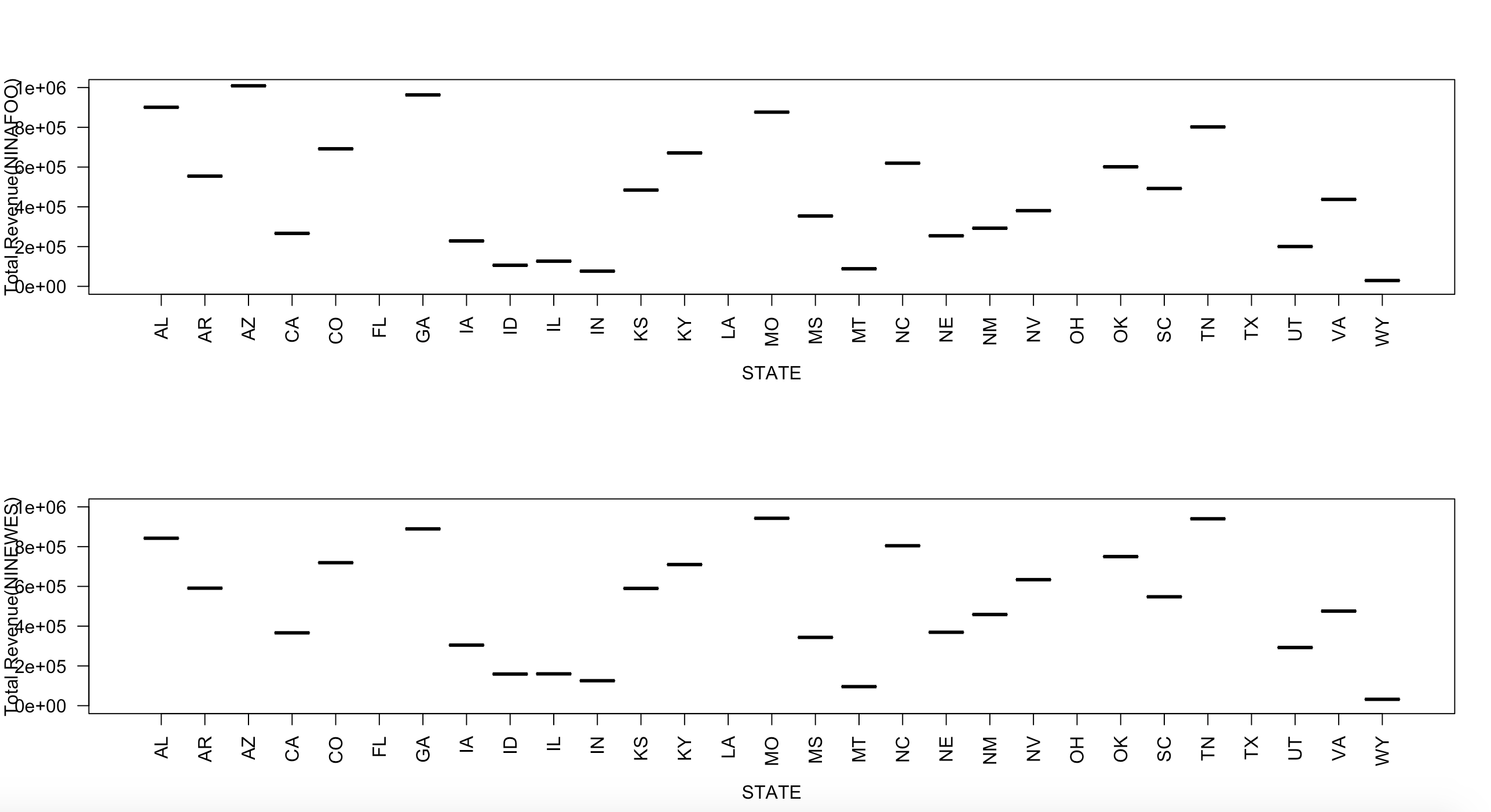
xlab = "STATE", ylab = "Revenue per Store(NINEWES)",ylim = c(0,60),

las =2)

图片包含 鸟, 天空, 群

描述已自动生成

Box plot for number of stores



Boxplot for Total revenue

图片包含 屏幕截图

描述已自动生成

Boxplot for Revenue per Store

Provide comments about visualizations.

The map vividly demonstrates the state with highest or lowest revenue and number of stores in different states.

Meanwhile, the boxplots can better compare the quantity difference of revenue, number of stores and even the revenue per store in different states, which is flexible but not so directly visualization.

1. ( 5 pts) Reflective statement

In the project, I get the required data in the Dillard’s dataset and visualize it with maps and plots. For further analysis, more visualization tools could be applied to the data and even make marketing strategy plans in different states.