01\_select\_Toronto

SPSS

2018-12-01

library(data.table)

user <- fread("data/yelp\_academic\_dataset\_user.csv")  
summary(user)

## user\_id name review\_count   
## Length:1518169 Length:1518169 Min. : 0.00   
## Class :character Class :character 1st Qu.: 2.00   
## Mode :character Mode :character Median : 5.00   
## Mean : 22.43   
## 3rd Qu.: 15.00   
## Max. :12723.00   
## yelping\_since friends useful   
## Length:1518169 Length:1518169 Min. : 0.00   
## Class :character Class :character 1st Qu.: 0.00   
## Mode :character Mode :character Median : 0.00   
## Mean : 28.67   
## 3rd Qu.: 3.00   
## Max. :258479.00   
## funny cool fans   
## Min. : 0.00 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 0.00 Median : 0.00 Median : 0.000   
## Mean : 13.99 Mean : 19.12 Mean : 1.454   
## 3rd Qu.: 0.00 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :242120.00 Max. :255909.00 Max. :8665.000   
## elite average\_stars compliment\_hot compliment\_more   
## Length:1518169 Min. :0.000 Min. : 0.00 Min. : 0.00   
## Class :character 1st Qu.:3.070 1st Qu.: 0.00 1st Qu.: 0.00   
## Mode :character Median :3.920 Median : 0.00 Median : 0.00   
## Mean :3.717 Mean : 2.35 Mean : 0.35   
## 3rd Qu.:4.670 3rd Qu.: 0.00 3rd Qu.: 0.00   
## Max. :5.000 Max. :33534.00 Max. :13494.00   
## compliment\_profile compliment\_cute compliment\_list   
## Min. : 0.000 Min. : 0.000 Min. : 0.00   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.00   
## Median : 0.000 Median : 0.000 Median : 0.00   
## Mean : 0.226 Mean : 0.203 Mean : 0.09   
## 3rd Qu.: 0.000 3rd Qu.: 0.000 3rd Qu.: 0.00   
## Max. :14165.000 Max. :13648.000 Max. :12663.00   
## compliment\_note compliment\_plain compliment\_cool   
## Min. : 0.00 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 0.00 Median : 0.00 Median : 0.000   
## Mean : 1.55 Mean : 3.24 Mean : 3.258   
## 3rd Qu.: 0.00 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :57475.00 Max. :48628.00 Max. :30699.000   
## compliment\_funny compliment\_writer compliment\_photos   
## Min. : 0.000 Min. : 0.000 Min. : 0.00   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.00   
## Median : 0.000 Median : 0.000 Median : 0.00   
## Mean : 3.258 Mean : 1.225 Mean : 1.16   
## 3rd Qu.: 0.000 3rd Qu.: 0.000 3rd Qu.: 0.00   
## Max. :30699.000 Max. :15433.000 Max. :82593.00

head(user)

## user\_id name review\_count yelping\_since friends  
## 1: lzlZwIpuSWXEnNS91wxjHw Susan 1 2015-09-28 None  
## 2: XvLBr-9smbI0m\_a7dXtB7w Daipayan 2 2015-09-05 None  
## 3: QPT4Ud4H5sJVr68yXhoWFw Andy 1 2016-07-21 None  
## 4: i5YitlHZpf0B3R0s\_8NVuw Jonathan 19 2014-08-04 None  
## 5: s4FoIXE\_LSGviTHBe8dmcg Shashank 3 2017-06-18 None  
## 6: ZcsZdHLiJGVvDHVjeTYYnQ Stacey X Joe 3 2014-08-27 None  
## useful funny cool fans elite average\_stars compliment\_hot  
## 1: 0 0 0 0 None 2.00 0  
## 2: 0 0 0 0 None 5.00 0  
## 3: 0 0 0 0 None 4.00 0  
## 4: 0 0 0 0 None 4.05 0  
## 5: 0 0 0 0 None 3.00 0  
## 6: 0 0 0 0 None 5.00 0  
## compliment\_more compliment\_profile compliment\_cute compliment\_list  
## 1: 0 0 0 0  
## 2: 0 0 0 0  
## 3: 0 0 0 0  
## 4: 0 0 0 0  
## 5: 0 0 0 0  
## 6: 0 0 0 0  
## compliment\_note compliment\_plain compliment\_cool compliment\_funny  
## 1: 0 0 0 0  
## 2: 0 0 0 0  
## 3: 0 0 0 0  
## 4: 0 0 0 0  
## 5: 0 0 0 0  
## 6: 0 0 0 0  
## compliment\_writer compliment\_photos  
## 1: 0 0  
## 2: 0 0  
## 3: 0 0  
## 4: 0 0  
## 5: 0 0  
## 6: 0 0

checkin <- fread("data/yelp\_academic\_dataset\_checkin.csv")  
summary(user)

## user\_id name review\_count   
## Length:1518169 Length:1518169 Min. : 0.00   
## Class :character Class :character 1st Qu.: 2.00   
## Mode :character Mode :character Median : 5.00   
## Mean : 22.43   
## 3rd Qu.: 15.00   
## Max. :12723.00   
## yelping\_since friends useful   
## Length:1518169 Length:1518169 Min. : 0.00   
## Class :character Class :character 1st Qu.: 0.00   
## Mode :character Mode :character Median : 0.00   
## Mean : 28.67   
## 3rd Qu.: 3.00   
## Max. :258479.00   
## funny cool fans   
## Min. : 0.00 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 0.00 Median : 0.00 Median : 0.000   
## Mean : 13.99 Mean : 19.12 Mean : 1.454   
## 3rd Qu.: 0.00 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :242120.00 Max. :255909.00 Max. :8665.000   
## elite average\_stars compliment\_hot compliment\_more   
## Length:1518169 Min. :0.000 Min. : 0.00 Min. : 0.00   
## Class :character 1st Qu.:3.070 1st Qu.: 0.00 1st Qu.: 0.00   
## Mode :character Median :3.920 Median : 0.00 Median : 0.00   
## Mean :3.717 Mean : 2.35 Mean : 0.35   
## 3rd Qu.:4.670 3rd Qu.: 0.00 3rd Qu.: 0.00   
## Max. :5.000 Max. :33534.00 Max. :13494.00   
## compliment\_profile compliment\_cute compliment\_list   
## Min. : 0.000 Min. : 0.000 Min. : 0.00   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.00   
## Median : 0.000 Median : 0.000 Median : 0.00   
## Mean : 0.226 Mean : 0.203 Mean : 0.09   
## 3rd Qu.: 0.000 3rd Qu.: 0.000 3rd Qu.: 0.00   
## Max. :14165.000 Max. :13648.000 Max. :12663.00   
## compliment\_note compliment\_plain compliment\_cool   
## Min. : 0.00 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 0.00 Median : 0.00 Median : 0.000   
## Mean : 1.55 Mean : 3.24 Mean : 3.258   
## 3rd Qu.: 0.00 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :57475.00 Max. :48628.00 Max. :30699.000   
## compliment\_funny compliment\_writer compliment\_photos   
## Min. : 0.000 Min. : 0.000 Min. : 0.00   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.00   
## Median : 0.000 Median : 0.000 Median : 0.00   
## Mean : 3.258 Mean : 1.225 Mean : 1.16   
## 3rd Qu.: 0.000 3rd Qu.: 0.000 3rd Qu.: 0.00   
## Max. :30699.000 Max. :15433.000 Max. :82593.00

head(checkin)

## time  
## 1: {'Fri-0': 2, 'Sat-0': 1, 'Sun-0': 1, 'Wed-0': 2, 'Fri-1': 1, 'Sat-1': 3, 'Thu-1': 1, 'Wed-1': 1, 'Sat-2': 1, 'Sun-2': 2, 'Thu-2': 1, 'Wed-2': 1, 'Fri-3': 1, 'Sun-3': 3, 'Mon-4': 1, 'Thu-4': 1, 'Tue-4': 2, 'Wed-4': 2, 'Sun-6': 1, 'Wed-6': 1, 'Thu-7': 1, 'Fri-10': 3, 'Mon-10': 1, 'Sat-10': 3, 'Sun-10': 3, 'Tue-10': 2, 'Mon-11': 1, 'Thu-11': 1, 'Wed-11': 2, 'Mon-12': 1, 'Sat-12': 1, 'Tue-12': 1, 'Sat-13': 3, 'Thu-13': 1, 'Tue-13': 2, 'Wed-13': 3, 'Fri-14': 2, 'Mon-14': 1, 'Sat-14': 1, 'Sun-14': 1, 'Tue-14': 1, 'Wed-14': 2, 'Fri-15': 1, 'Mon-15': 1, 'Sat-15': 1, 'Thu-15': 2, 'Tue-15': 1, 'Fri-16': 1, 'Sat-16': 2, 'Sun-16': 1, 'Thu-16': 1, 'Tue-16': 1, 'Fri-17': 1, 'Mon-17': 3, 'Sat-17': 4, 'Sun-17': 1, 'Tue-17': 1, 'Wed-17': 1, 'Fri-18': 1, 'Mon-18': 1, 'Sat-18': 4, 'Sun-18': 1, 'Thu-18': 1, 'Tue-18': 2, 'Wed-18': 1, 'Fri-19': 3, 'Mon-19': 1, 'Sun-19': 2, 'Thu-19': 1, 'Wed-19': 1, 'Mon-20': 1, 'Sun-20': 5, 'Thu-20': 4, 'Tue-20': 2, 'Wed-20': 3, 'Fri-21': 4, 'Mon-21': 1, 'Sat-21': 1, 'Sun-21': 1, 'Thu-21': 4, 'Tue-21': 1, 'Wed-21': 1, 'Fri-22': 3, 'Thu-22': 1, 'Fri-23': 1, 'Mon-23': 2, 'Sat-23': 3, 'Sun-23': 1, 'Thu-23': 2, 'Tue-23': 2}  
## 2: {'Mon-13': 1, 'Thu-13': 1, 'Sat-16': 1, 'Wed-17': 1, 'Sun-19': 1, 'Thu-20': 1, 'Sat-21': 1}  
## 3: {'Thu-0': 1, 'Mon-1': 1, 'Mon-12': 1, 'Sat-16': 1, 'Sun-22': 1, 'Fri-23': 1}  
## 4: {'Sun-14': 1, 'Fri-18': 1, 'Mon-20': 1}  
## 5: {'Sat-3': 1, 'Sun-18': 1, 'Sat-21': 1, 'Sat-23': 1, 'Thu-23': 1}  
## 6: {'Sat-17': 1, 'Sat-18': 1}  
## business\_id  
## 1: 7KPBkxAOEtb3QeIL9PEErg  
## 2: kREVIrSBbtqBhIYkTccQUg  
## 3: tJRDll5yqpZwehenzE2cSg  
## 4: tZccfdl6JNw-j5BKnCTIQQ  
## 5: r1p7RAMzCV\_6NPF0dNoR3g  
## 6: OVkBMWxUBXGoO4K3\_CeJ7g

review <- fread("data/yelp\_academic\_dataset\_review.csv")  
summary(user)

## user\_id name review\_count   
## Length:1518169 Length:1518169 Min. : 0.00   
## Class :character Class :character 1st Qu.: 2.00   
## Mode :character Mode :character Median : 5.00   
## Mean : 22.43   
## 3rd Qu.: 15.00   
## Max. :12723.00   
## yelping\_since friends useful   
## Length:1518169 Length:1518169 Min. : 0.00   
## Class :character Class :character 1st Qu.: 0.00   
## Mode :character Mode :character Median : 0.00   
## Mean : 28.67   
## 3rd Qu.: 3.00   
## Max. :258479.00   
## funny cool fans   
## Min. : 0.00 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 0.00 Median : 0.00 Median : 0.000   
## Mean : 13.99 Mean : 19.12 Mean : 1.454   
## 3rd Qu.: 0.00 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :242120.00 Max. :255909.00 Max. :8665.000   
## elite average\_stars compliment\_hot compliment\_more   
## Length:1518169 Min. :0.000 Min. : 0.00 Min. : 0.00   
## Class :character 1st Qu.:3.070 1st Qu.: 0.00 1st Qu.: 0.00   
## Mode :character Median :3.920 Median : 0.00 Median : 0.00   
## Mean :3.717 Mean : 2.35 Mean : 0.35   
## 3rd Qu.:4.670 3rd Qu.: 0.00 3rd Qu.: 0.00   
## Max. :5.000 Max. :33534.00 Max. :13494.00   
## compliment\_profile compliment\_cute compliment\_list   
## Min. : 0.000 Min. : 0.000 Min. : 0.00   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.00   
## Median : 0.000 Median : 0.000 Median : 0.00   
## Mean : 0.226 Mean : 0.203 Mean : 0.09   
## 3rd Qu.: 0.000 3rd Qu.: 0.000 3rd Qu.: 0.00   
## Max. :14165.000 Max. :13648.000 Max. :12663.00   
## compliment\_note compliment\_plain compliment\_cool   
## Min. : 0.00 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000   
## Median : 0.00 Median : 0.00 Median : 0.000   
## Mean : 1.55 Mean : 3.24 Mean : 3.258   
## 3rd Qu.: 0.00 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :57475.00 Max. :48628.00 Max. :30699.000   
## compliment\_funny compliment\_writer compliment\_photos   
## Min. : 0.000 Min. : 0.000 Min. : 0.00   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.00   
## Median : 0.000 Median : 0.000 Median : 0.00   
## Mean : 3.258 Mean : 1.225 Mean : 1.16   
## 3rd Qu.: 0.000 3rd Qu.: 0.000 3rd Qu.: 0.00   
## Max. :30699.000 Max. :15433.000 Max. :82593.00

head(review)

## review\_id user\_id business\_id  
## 1: x7mDIiDB3jEiPGPHOmDzyw msQe1u7Z\_XuqjGoqhB0J5g iCQpiavjjPzJ5\_3gPD5Ebg  
## 2: dDl8zu1vWPdKGihJrwQbpw msQe1u7Z\_XuqjGoqhB0J5g pomGBqfbxcqPv14c3XH-ZQ  
## 3: LZp4UX5zK3e-c5ZGSeo3kA msQe1u7Z\_XuqjGoqhB0J5g jtQARsP6P-LbkyjbO1qNGg  
## 4: Er4NBWCmCD4nM8\_p1GRdow msQe1u7Z\_XuqjGoqhB0J5g elqbBhBfElMNSrjFqW3now  
## 5: jsDu6QEJHbwP2Blom1PLCA msQe1u7Z\_XuqjGoqhB0J5g Ums3gaP2qM3W1XcA5r6SsQ  
## 6: pfavA0hr3nyqO61oupj-lA msQe1u7Z\_XuqjGoqhB0J5g vgfcTvK81oD4r50NMjU2Ag  
## stars date  
## 1: 2 2011-02-25  
## 2: 5 2012-11-13  
## 3: 1 2014-10-23  
## 4: 2 2011-02-25  
## 5: 5 2014-09-05  
## 6: 1 2011-02-25  
## text  
## 1: The pizza was okay. Not the best I've had. I prefer Biaggio's on Flamingo / Fort Apache. The chef there can make a MUCH better NY style pizza. The pizzeria @ Cosmo was over priced for the quality and lack of personality in the food. Biaggio's is a much better pick if youre going for italian - family owned, home made recipes, people that actually CARE if you like their food. You dont get that at a pizzeria in a casino. I dont care what you say...  
## 2: I love this place! My fiance And I go here atleast once a week. The portions are huge! Food is amazing. I love their carne asada. They have great lunch specials... Leticia is super nice and cares about what you think of her restaurant. You have to try their cheese enchiladas too the sauce is different And amazing!!!  
## 3: Terrible. Dry corn bread. Rib tips were all fat and mushy and had no flavor. If you want bbq in this neighborhood go to john mulls roadkill grill. Trust me.  
## 4: Back in 2005-2007 this place was my FAVORITE thai place EVER. I'd go here ALLLLL the time. I never had any complaints. Once they started to get more known and got busy, their service started to suck and their portion sizes got cut in half. I have a huge problem with paying MORE for way less food. The last time I went there I had the Pork Pad se Ew and it tasted good, but I finished my plate and was still hungry. I used to know the manager here and she would greet me with a ""Hello Melissa, nice to see you again, diet coke & pad thai or pad se ew?"" Now a days, I know she still knows me but she disregards my presence. Also, I had asked her what was up with the new portion sizes and she had no answer for me. Great food but not worth the money. I havent been back in over a year because I refuse to pay $10-15 for dinner and still be hungry after. Sorry PinKaow, you are not what you used to be!!  
## 5: Delicious healthy food. The steak is amazing. Fish and pork are awesome too. Service is above and beyond. Not a bad thing to say about this place. Worth every penny!  
## 6: This place sucks. The customer service is horrible. They dont serve food unless you order a pizza from a neighboring restaurant. Who does that? They dont control their crowd. Many times I've gone I've seen fights. The bartenders suck - I've almost got in a fight with one because she was a complete bitch. Refused to serve me a drink because she was ""busy"" celebrating her friends birthday BEHIND THE BAR. This place is ridiculous. I will NEVER go there again.. EVER.  
## useful funny cool  
## 1: 0 0 0  
## 2: 0 0 0  
## 3: 3 1 1  
## 4: 2 0 0  
## 5: 0 0 0  
## 6: 2 0 0

business <- fread("data/yelp\_academic\_dataset\_business.csv")  
summary(business)

## business\_id name neighborhood   
## Length:188593 Length:188593 Length:188593   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## address city state   
## Length:188593 Length:188593 Length:188593   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## postal\_code latitude longitude stars   
## Length:188593 Min. :-71.75 Min. :-180.00 Min. :1.000   
## Class :character 1st Qu.: 33.63 1st Qu.:-112.28 1st Qu.:3.000   
## Mode :character Median : 36.14 Median :-111.78 Median :3.500   
## Mean : 38.51 Mean : -97.49 Mean :3.632   
## 3rd Qu.: 43.59 3rd Qu.: -79.98 3rd Qu.:4.500   
## Max. : 85.05 Max. : 115.09 Max. :5.000   
## NA's :6 NA's :6   
## review\_count is\_open attributes categories   
## Min. : 3.0 Min. :0.0000 Length:188593 Length:188593   
## 1st Qu.: 4.0 1st Qu.:1.0000 Class :character Class :character   
## Median : 9.0 Median :1.0000 Mode :character Mode :character   
## Mean : 31.8 Mean :0.8304   
## 3rd Qu.: 24.0 3rd Qu.:1.0000   
## Max. :7968.0 Max. :1.0000   
##   
## hours   
## Length:188593   
## Class :character   
## Mode :character   
##   
##   
##   
##

head(business)

## business\_id name  
## 1: Apn5Q\_b6Nz61Tq4XzPdf9A Minhas Micro Brewery  
## 2: AjEbIBw6ZFfln7ePHha9PA CK'S BBQ & Catering  
## 3: O8S5hYJ1SMc8fA4QBtVujA La Bastringue  
## 4: bFzdJJ3wp3PZssNEsyU23g Geico Insurance  
## 5: 8USyCYqpScwiNEb58Bt6CA Action Engine  
## 6: 45bWSZtniwPRiqlivpS8Og The Coffee Bean & Tea Leaf  
## neighborhood address city  
## 1: 1314 44 Avenue NE Calgary  
## 2: Henderson  
## 3: Rosemont-La Petite-Patrie 1335 rue Beaubien E Montréal  
## 4: 211 W Monroe St Phoenix  
## 5: 2005 Alyth Place SE Calgary  
## 6: 20235 N Cave Creek Rd, Ste 1115 Phoenix  
## state postal\_code latitude longitude stars review\_count is\_open  
## 1: AB T2E 6L6 51.09181 -114.0317 4.0 24 1  
## 2: NV 89002 35.96073 -114.9398 4.5 3 0  
## 3: QC H2G 1K7 45.54050 -73.5993 4.0 5 0  
## 4: AZ 85003 33.45000 -112.0770 1.5 8 1  
## 5: AB T2H 0N5 51.03559 -114.0274 2.0 4 1  
## 6: AZ 85024 33.67138 -112.0300 4.0 63 1  
## attributes  
## 1: {'BikeParking': 'False', 'BusinessAcceptsCreditCards': 'True', 'BusinessParking': ""{'garage': False, 'street': True, 'validated': False, 'lot': False, 'valet': False}"", 'GoodForKids': 'True', 'HasTV': 'True', 'NoiseLevel': 'average', 'OutdoorSeating': 'False', 'RestaurantsAttire': 'casual', 'RestaurantsDelivery': 'False', 'RestaurantsGoodForGroups': 'True', 'RestaurantsPriceRange2': '2', 'RestaurantsReservations': 'True', 'RestaurantsTakeOut': 'True'}  
## 2: {'Alcohol': 'none', 'BikeParking': 'False', 'BusinessAcceptsCreditCards': 'True', 'BusinessParking': ""{'garage': False, 'street': True, 'validated': False, 'lot': True, 'valet': False}"", 'Caters': 'True', 'DogsAllowed': 'True', 'DriveThru': 'False', 'GoodForKids': 'True', 'GoodForMeal': ""{'dessert': False, 'latenight': False, 'lunch': False, 'dinner': False, 'breakfast': False, 'brunch': False}"", 'HasTV': 'False', 'OutdoorSeating': 'True', 'RestaurantsAttire': 'casual', 'RestaurantsDelivery': 'False', 'RestaurantsGoodForGroups': 'True', 'RestaurantsPriceRange2': '2', 'RestaurantsReservations': 'False', 'RestaurantsTableService': 'False', 'RestaurantsTakeOut': 'True', 'WheelchairAccessible': 'True', 'WiFi': 'no'}  
## 3: {'Alcohol': 'beer\_and\_wine', 'Ambience': ""{'romantic': False, 'intimate': False, 'classy': False, 'hipster': False, 'touristy': False, 'trendy': False, 'upscale': False, 'casual': False}"", 'BikeParking': 'True', 'BusinessAcceptsCreditCards': 'False', 'BusinessParking': ""{'garage': False, 'street': False, 'validated': False, 'lot': False, 'valet': False}"", 'Caters': 'False', 'GoodForKids': 'True', 'GoodForMeal': ""{'dessert': False, 'latenight': False, 'lunch': False, 'dinner': False, 'breakfast': False, 'brunch': False}"", 'HasTV': 'True', 'NoiseLevel': 'average', 'OutdoorSeating': 'False', 'RestaurantsAttire': 'casual', 'RestaurantsDelivery': 'False', 'RestaurantsGoodForGroups': 'True', 'RestaurantsPriceRange2': '2', 'RestaurantsReservations': 'True', 'RestaurantsTableService': 'True', 'RestaurantsTakeOut': 'False', 'WiFi': 'free'}  
## 4:   
## 5: {'BusinessAcceptsCreditCards': 'True'}  
## 6: {'BikeParking': 'True', 'BusinessAcceptsCreditCards': 'True', 'BusinessParking': ""{'garage': False, 'street': False, 'validated': False, 'lot': True, 'valet': False}"", 'Caters': 'False', 'OutdoorSeating': 'True', 'RestaurantsPriceRange2': '1', 'RestaurantsTakeOut': 'True', 'WheelchairAccessible': 'True', 'WiFi': 'free'}  
## categories  
## 1: Tours, Breweries, Pizza, Restaurants, Food, Hotels & Travel  
## 2: Chicken Wings, Burgers, Caterers, Street Vendors, Barbeque, Food Trucks, Food, Restaurants, Event Planning & Services  
## 3: Breakfast & Brunch, Restaurants, French, Sandwiches, Cafes  
## 4: Insurance, Financial Services  
## 5: Home & Garden, Nurseries & Gardening, Shopping, Local Services, Automotive, Electronics Repair  
## 6: Coffee & Tea, Food  
## hours  
## 1: {'Monday': '8:30-17:0', 'Tuesday': '11:0-21:0', 'Wednesday': '11:0-21:0', 'Thursday': '11:0-21:0', 'Friday': '11:0-21:0', 'Saturday': '11:0-21:0'}  
## 2: {'Friday': '17:0-23:0', 'Saturday': '17:0-23:0', 'Sunday': '17:0-23:0'}  
## 3: {'Monday': '10:0-22:0', 'Tuesday': '10:0-22:0', 'Wednesday': '10:0-22:0', 'Thursday': '10:0-22:0', 'Friday': '10:0-22:0', 'Saturday': '10:0-22:0', 'Sunday': '10:0-22:0'}  
## 4:   
## 5: {'Monday': '8:0-17:0', 'Tuesday': '8:0-17:0', 'Wednesday': '8:0-17:0', 'Thursday': '8:0-17:0', 'Friday': '8:0-17:0'}  
## 6: {'Monday': '5:30-20:0', 'Tuesday': '5:30-20:0', 'Wednesday': '5:30-20:0', 'Thursday': '5:30-20:0', 'Friday': '5:30-21:0', 'Saturday': '5:30-21:0', 'Sunday': '6:30-19:0'}

## select business in toronto  
torontoBusiness <- business[business$city == "Toronto"]  
torontoBusinessID <- torontoBusiness$business\_id

## get review of toronto business -> get user/author of these reviews  
  
torontoReview <- review[review$business\_id %in% torontoBusinessID]  
reviewInfo <- torontoReview[, c("user\_id", "business\_id", "date")]  
  
library(sqldf)

## Loading required package: gsubfn

## Loading required package: proto

## Warning in doTryCatch(return(expr), name, parentenv, handler): unable to load shared object '/Library/Frameworks/R.framework/Resources/modules//R\_X11.so':  
## dlopen(/Library/Frameworks/R.framework/Resources/modules//R\_X11.so, 6): Library not loaded: /opt/X11/lib/libSM.6.dylib  
## Referenced from: /Library/Frameworks/R.framework/Resources/modules//R\_X11.so  
## Reason: image not found

## Could not load tcltk. Will use slower R code instead.

## Loading required package: RSQLite

user\_freq <- sqldf("select \*, count(\*) as freq, max(date) as max, min(date) as min from reviewInfo group by user\_id order by freq desc")  
user\_freq["span"] <- as.numeric(as.Date(user\_freq$max) - as.Date(user\_freq$min))  
summary(user\_freq)

## user\_id business\_id date   
## Length:103262 Length:103262 Length:103262   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
## freq max min span   
## Min. : 1.000 Length:103262 Length:103262 Min. : 0.0   
## 1st Qu.: 1.000 Class :character Class :character 1st Qu.: 0.0   
## Median : 1.000 Mode :character Mode :character Median : 0.0   
## Mean : 4.598 Mean : 254.8   
## 3rd Qu.: 3.000 3rd Qu.: 251.0   
## Max. :1564.000 Max. :3505.0

# select users that have been in toronto for more than one month, otherwise he/she may be a tourist or inactive  
activeUser <- user\_freq[user\_freq$span>30, ]  
torontoUserID <- activeUser$user\_id  
torontoUser <- user[user$user\_id %in% torontoUserID]  
  
torontoUser <- sqldf("select \* from torontoUser left join activeUser on torontouser.user\_id = activeUser.user\_id")  
  
torontoReview <- torontoReview[(torontoReview$user\_id %in% torontoUserID) & (torontoReview$business\_id %in% torontoBusinessID), ]

library(stringr)  
torontoUser$friend\_count <- str\_count(torontoUser$friends, ",")  
torontoUser$friends <- NULL  
torontoUser$user\_id..23 <- NULL  
torontoUser$business\_id <- NULL

summary(torontoUser)

## user\_id name review\_count   
## Length:35458 Length:35458 Min. : 0.00   
## Class :character Class :character 1st Qu.: 4.00   
## Mode :character Mode :character Median : 9.00   
## Mean : 39.58   
## 3rd Qu.: 23.00   
## Max. :11942.00   
## yelping\_since useful funny   
## Length:35458 Min. : 0.00 Min. : 0.00   
## Class :character 1st Qu.: 0.00 1st Qu.: 0.00   
## Mode :character Median : 1.00 Median : 0.00   
## Mean : 66.36 Mean : 29.27   
## 3rd Qu.: 5.00 3rd Qu.: 1.00   
## Max. :89558.00 Max. :57867.00   
## cool fans elite average\_stars   
## Min. : 0.00 Min. : 0.000 Length:35458 Min. :1.000   
## 1st Qu.: 0.00 1st Qu.: 0.000 Class :character 1st Qu.:3.100   
## Median : 0.00 Median : 0.000 Mode :character Median :3.660   
## Mean : 46.27 Mean : 2.978 Mean :3.547   
## 3rd Qu.: 1.00 3rd Qu.: 1.000 3rd Qu.:4.070   
## Max. :71676.00 Max. :2311.000 Max. :5.000   
## compliment\_hot compliment\_more compliment\_profile   
## Min. : 0.000 Min. : 0.0000 Min. : 0.0000   
## 1st Qu.: 0.000 1st Qu.: 0.0000 1st Qu.: 0.0000   
## Median : 0.000 Median : 0.0000 Median : 0.0000   
## Mean : 4.357 Mean : 0.6888 Mean : 0.4505   
## 3rd Qu.: 0.000 3rd Qu.: 0.0000 3rd Qu.: 0.0000   
## Max. :11376.000 Max. :833.0000 Max. :1421.0000   
## compliment\_cute compliment\_list compliment\_note   
## Min. : 0.0000 Min. : 0.0000 Min. : 0.000   
## 1st Qu.: 0.0000 1st Qu.: 0.0000 1st Qu.: 0.000   
## Median : 0.0000 Median : 0.0000 Median : 0.000   
## Mean : 0.3437 Mean : 0.1895 Mean : 3.513   
## 3rd Qu.: 0.0000 3rd Qu.: 0.0000 3rd Qu.: 0.000   
## Max. :607.0000 Max. :327.0000 Max. :3194.000   
## compliment\_plain compliment\_cool compliment\_funny   
## Min. : 0.000 Min. : 0.000 Min. : 0.000   
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.000   
## Median : 0.000 Median : 0.000 Median : 0.000   
## Mean : 8.247 Mean : 6.822 Mean : 6.822   
## 3rd Qu.: 0.000 3rd Qu.: 0.000 3rd Qu.: 0.000   
## Max. :11579.000 Max. :13014.000 Max. :13014.000   
## compliment\_writer compliment\_photos date   
## Min. : 0.00 Min. : 0.000 Length:35458   
## 1st Qu.: 0.00 1st Qu.: 0.000 Class :character   
## Median : 0.00 Median : 0.000 Mode :character   
## Mean : 2.81 Mean : 2.252   
## 3rd Qu.: 0.00 3rd Qu.: 0.000   
## Max. :5668.00 Max. :11279.000   
## freq max min span   
## Min. : 2.00 Length:35458 Length:35458 Min. : 31.0   
## 1st Qu.: 3.00 Class :character Class :character 1st Qu.: 227.0   
## Median : 4.00 Mode :character Mode :character Median : 556.0   
## Mean : 10.61 Mean : 740.4   
## 3rd Qu.: 9.00 3rd Qu.:1075.0   
## Max. :1564.00 Max. :3505.0   
## friend\_count   
## Min. : 0.00   
## 1st Qu.: 0.00   
## Median : 1.00   
## Mean : 38.47   
## 3rd Qu.: 28.00   
## Max. :12197.00

summary(torontoBusiness)

## business\_id name neighborhood   
## Length:18233 Length:18233 Length:18233   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## address city state   
## Length:18233 Length:18233 Length:18233   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## postal\_code latitude longitude stars   
## Length:18233 Min. :-9.095 Min. :-80.62 Min. :1.00   
## Class :character 1st Qu.:43.651 1st Qu.:-79.42 1st Qu.:3.00   
## Mode :character Median :43.664 Median :-79.39 Median :3.50   
## Mean :43.675 Mean :-79.39 Mean :3.49   
## 3rd Qu.:43.688 3rd Qu.:-79.38 3rd Qu.:4.00   
## Max. :45.029 Max. :-51.83 Max. :5.00   
## NA's :1 NA's :1   
## review\_count is\_open attributes categories   
## Min. : 3.00 Min. :0.0000 Length:18233 Length:18233   
## 1st Qu.: 4.00 1st Qu.:1.0000 Class :character Class :character   
## Median : 9.00 Median :1.0000 Mode :character Mode :character   
## Mean : 26.04 Mean :0.7691   
## 3rd Qu.: 24.00 3rd Qu.:1.0000   
## Max. :1837.00 Max. :1.0000   
##   
## hours   
## Length:18233   
## Class :character   
## Mode :character   
##   
##   
##   
##

summary(torontoReview)

## review\_id user\_id business\_id   
## Length:376359 Length:376359 Length:376359   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
## stars date text   
## Length:376359 Length:376359 Length:376359   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
## useful funny cool   
## Min. : 0.000 Min. : 0.0000 Min. : 0.0000   
## 1st Qu.: 0.000 1st Qu.: 0.0000 1st Qu.: 0.0000   
## Median : 1.000 Median : 0.0000 Median : 0.0000   
## Mean : 1.519 Mean : 0.5111 Mean : 0.6333   
## 3rd Qu.: 2.000 3rd Qu.: 0.0000 3rd Qu.: 1.0000   
## Max. :272.000 Max. :193.0000 Max. :96.0000

write.csv(torontoUser, "toronto\_user.csv", row.names = FALSE)  
write.csv(torontoBusiness, "toronto\_business.csv", row.names = FALSE)  
write.csv(torontoReview, "toronto\_review.csv", row.names = FALSE)