Data Science Assignments

Wouter van Veelen

April 21, 2019

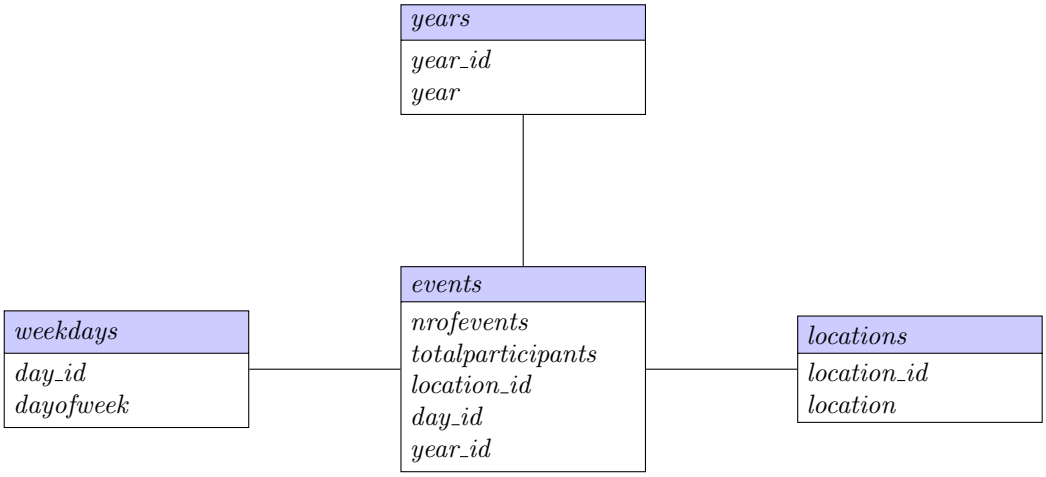
1 Assignment 1

1.1 a

1.1.1 1

nrofevents and *totalparticipants* are facts. *dayofweek*, *year* and *location* are dimensions.

1.1.2 2



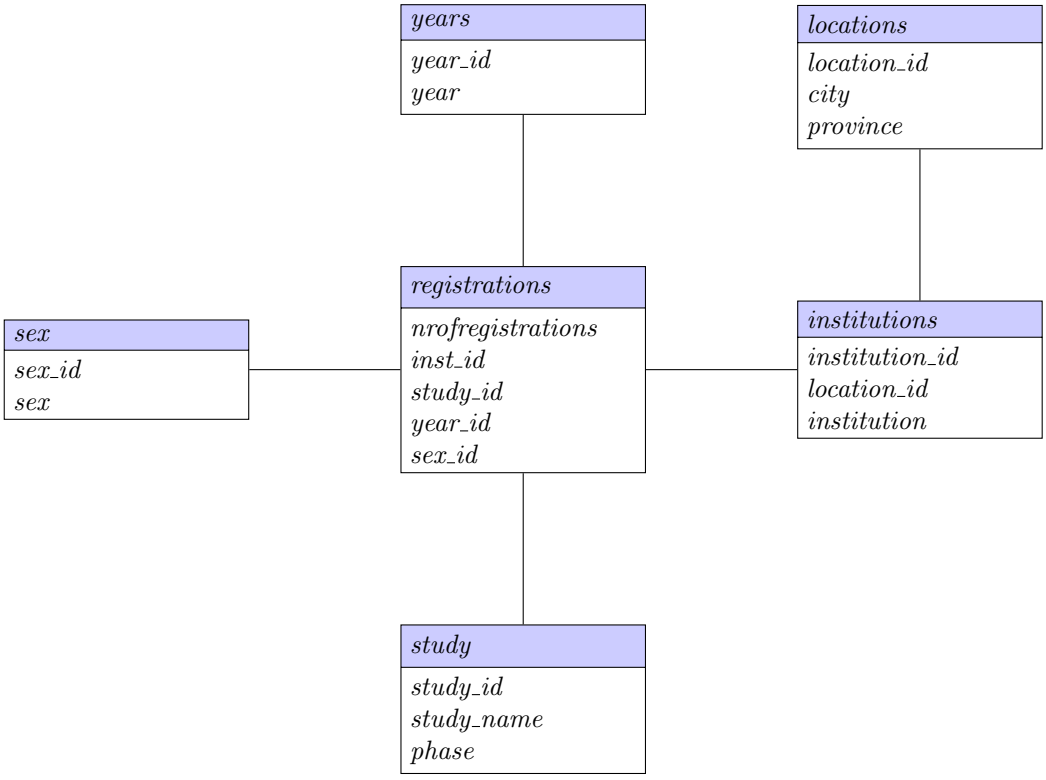
(Since all 'branches' all contain only a single entry, it is inefficiënt to make them all separate tables in the database. Locations is likely to be used separately, but the rest can be integrated in the main table.)

1.2 b

1.2.1 1

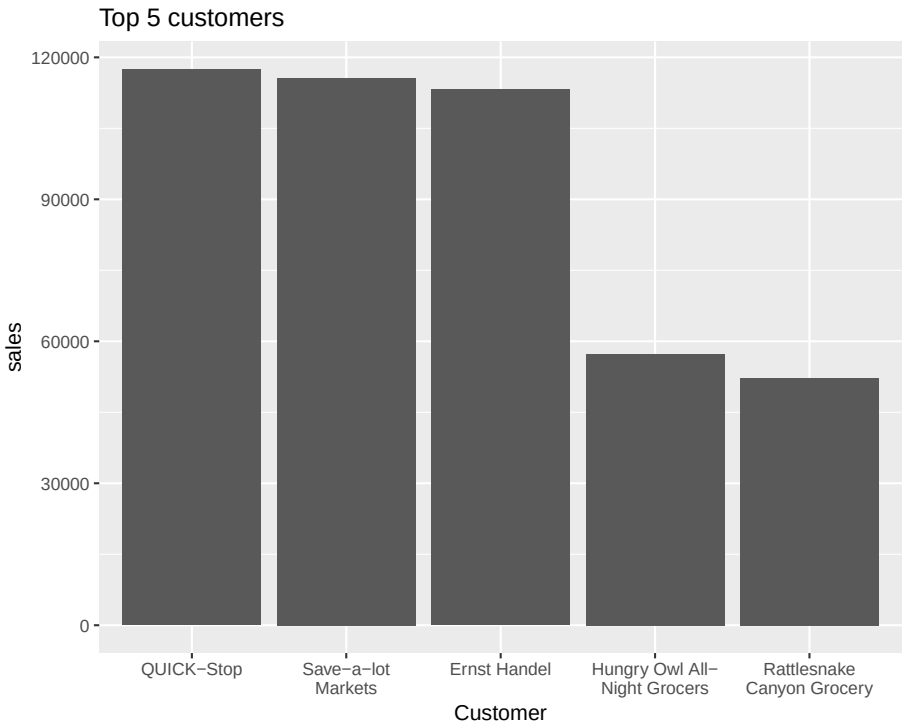
nrofregistrations is the only fact in the table, all of the others are dimensions.

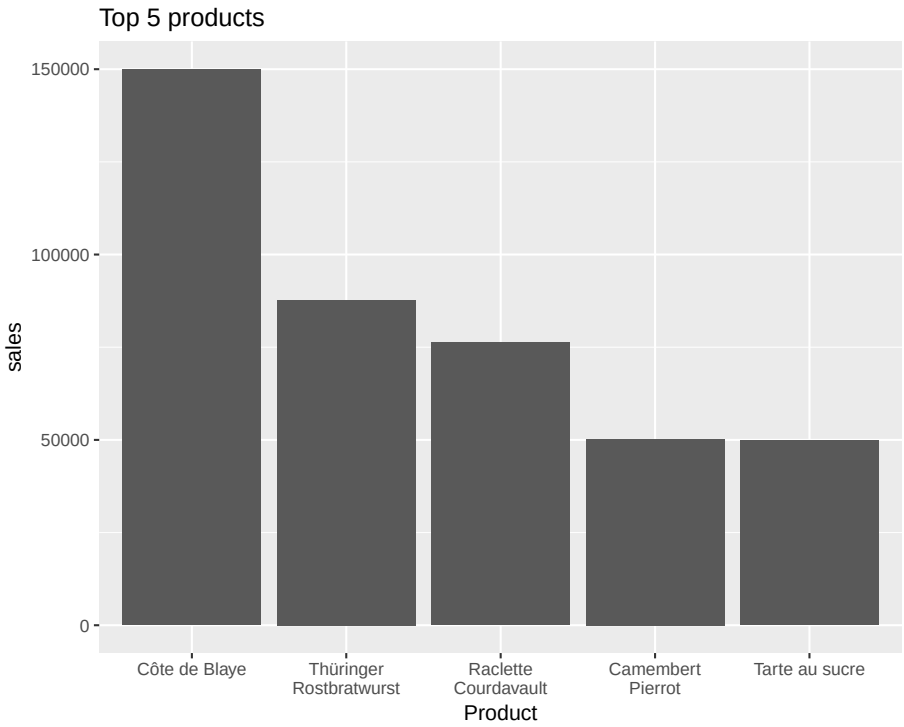
1.2.2 2



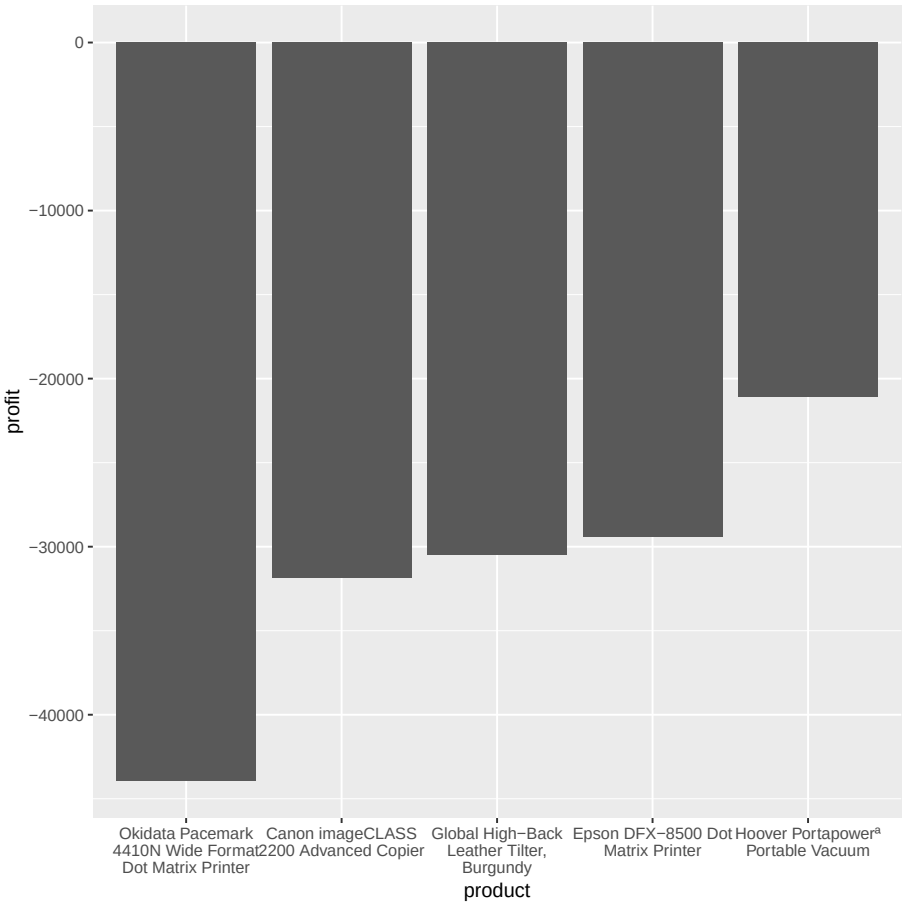
2 Assignment 2

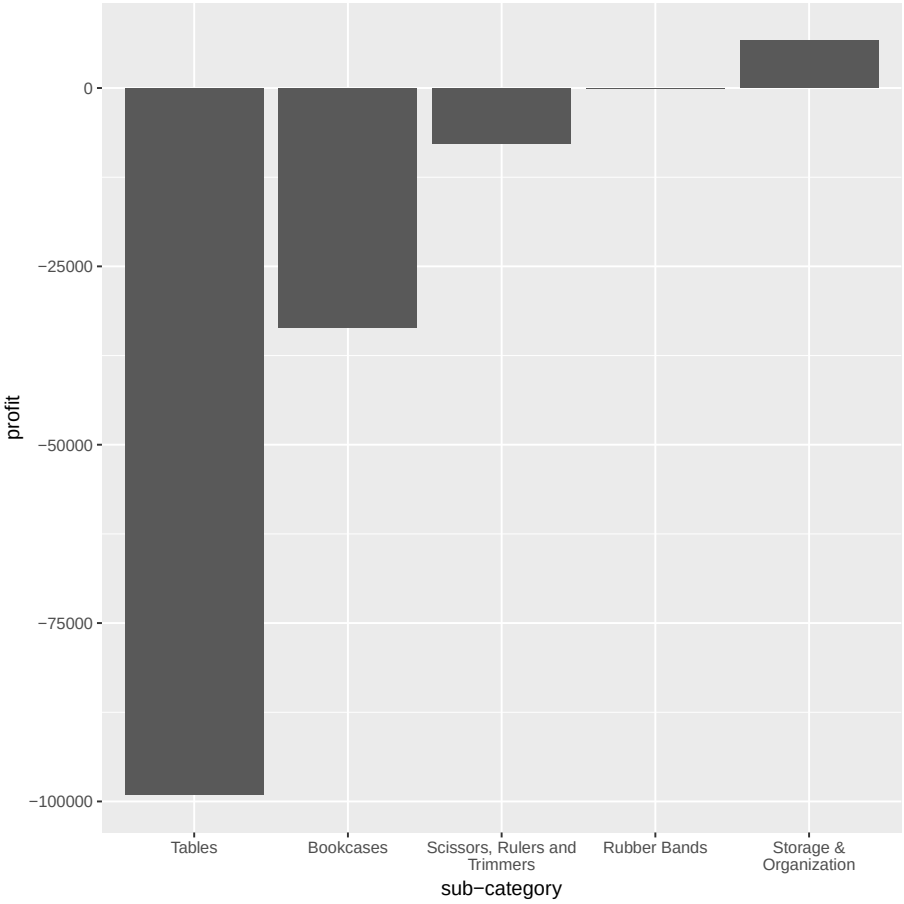
See the R script for the code, plots were made in ggplot

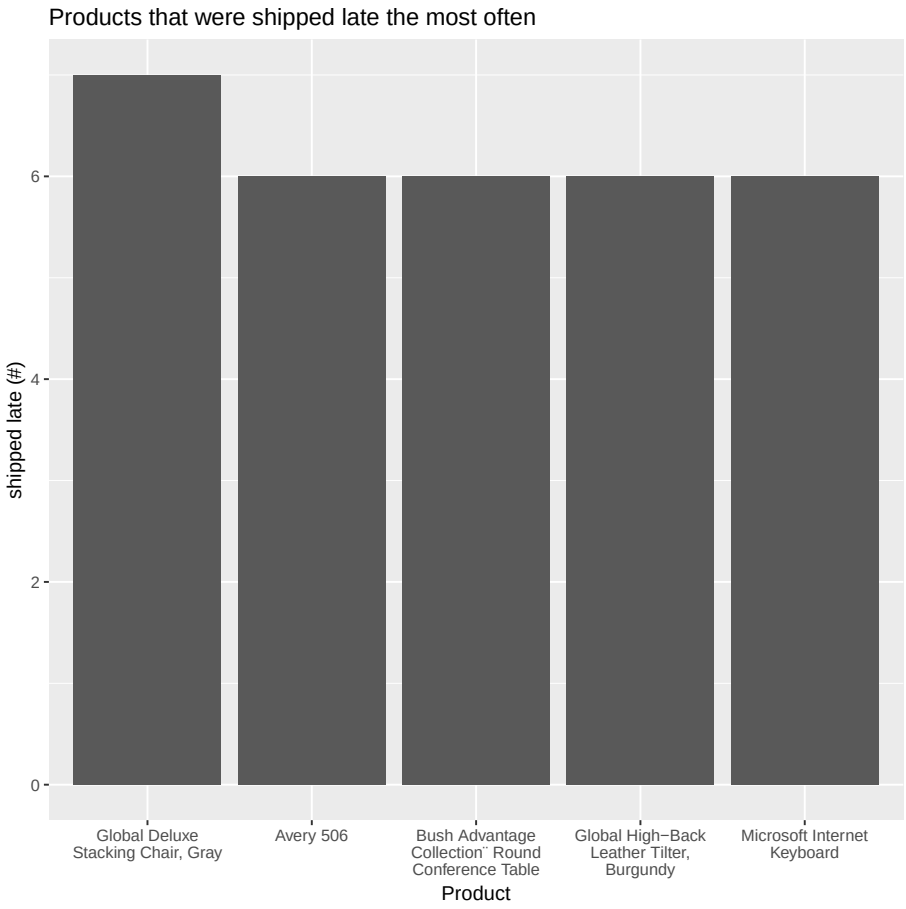

Top 5 customers
120000
90000
60000
30000
0
sales
QUICK−Stop
Save−a−lot Markets
Ernst Handel
Hungry Owl All−Night Grocers
Rattlesnake Canyon Grocery
Customer

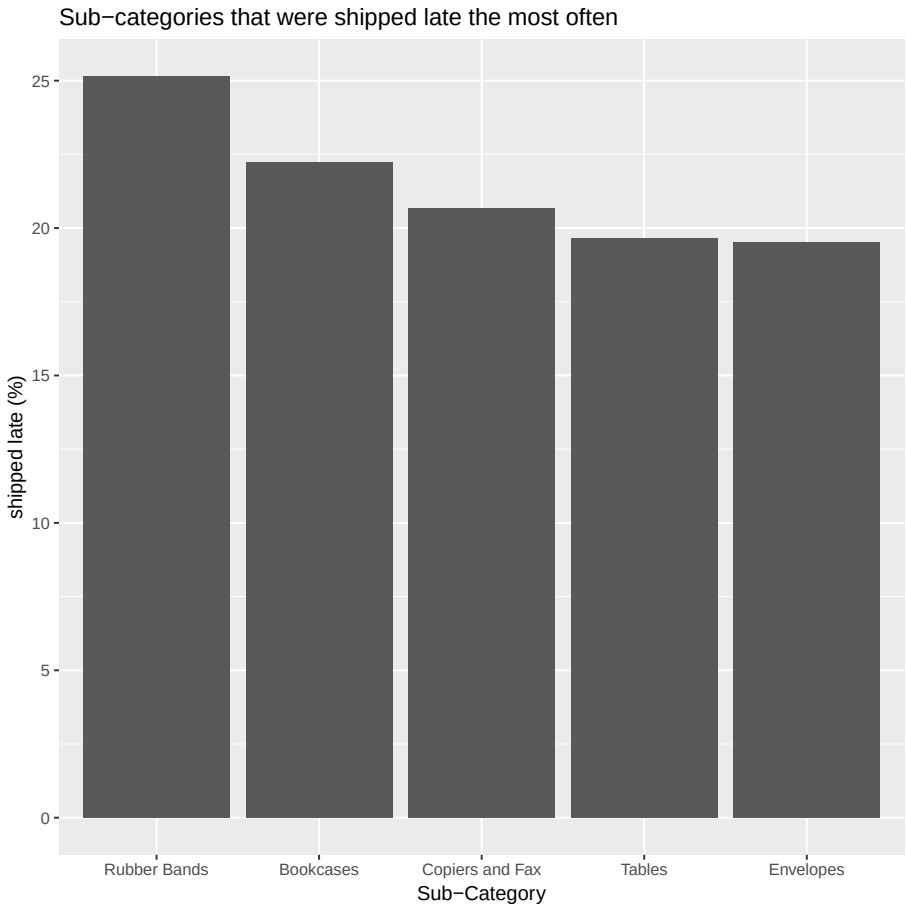

Top 5 products
150000
100000
50000
0
sales
Côte de Blaye
Thüringer Rostbratwurst
Raclette Courdavault
Camembert Pierrot
Tarte au sucre
Product

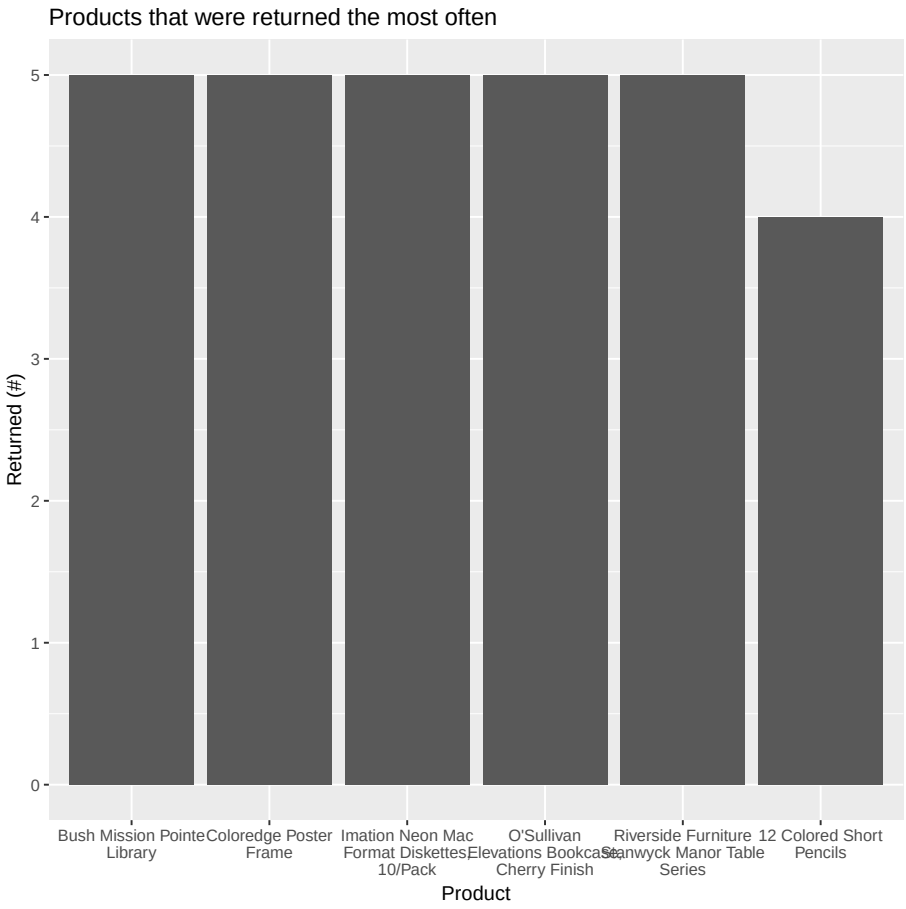
3 Assignment 3

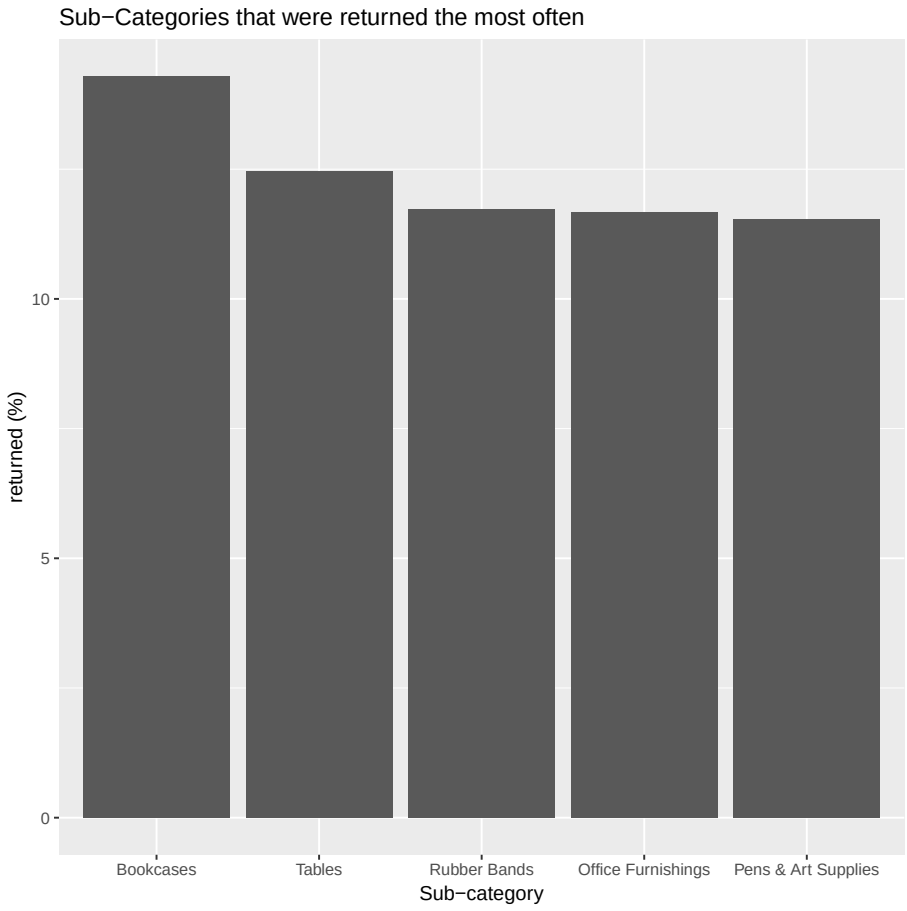



0
−25000
−50000
−75000
−100000
profit
Tables
Bookcases
Scissors, Rulers and Trimmers
Rubber Bands
Storage & Organization
sub−category


Products that were shipped late the most often
shipped late (#)
0
2
4
6
Global Deluxe Stacking Chair, Gray
Avery 506
Bush Advantage Collection¨ Round Conference Table
Global High−Back Leather Tilter, Burgundy
Microsoft Internet Keyboard
Product


Sub−categories that were shipped late the most often
shipped late (%)
25
20
15
10
5
0
Rubber Bands
Bookcases
Copiers and Fax
Tables
Envelopes
Sub−Category


Products that were returned the most often
Returned (#)
5
4
3
2
1
0
Bush Mission Pointe Library
Coloredge Poster Frame
Imation Neon Mac Format Diskettes 10/Pack
O'Sullivan Elevations Bookcase, Cherry Finish
Riverside Furniture Stanwyck Manor Table Series
12 Colored Short Pencils
Product



As for the queries:

```
Tables from Assignments 3:
> dbGetQuery(con,
+             "SELECT table_name FROM information_schema.tables
+              WHERE table_schema='ass3'")
  table_name
1      sales
2    product
3   customer



Tables from assignment 2:
> str(cust)
```

```
'data.frame':   89 obs. of  3 variables:
 $ name    : chr  "Alfreds Futterkiste" "Ana Trujillo Emparedados y helados" "Ant
 $ country: chr  "Germany" "Mexico" "Mexico" "UK" ...
 $ cust_id: int  1 2 3 4 5 6 7 8 9 10 ...
> str(sal)
'data.frame':   2155 obs. of  5 variables:
 $ order_id  : num  10248 10248 10248 10249 10249 ...
 $ order_date: chr  "9-11-2007" "9-11-2007" "9-11-2007" "10-11-2007" ...
 $ sales     : num  168 174 98 167 1696 ...
 $ productid : int  50 41 61 70 37 35 37 30 35 26 ...
 $ cust_id   : int  84 84 84 78 78 34 34 34 83 83 ...
> str(prod)
'data.frame':   77 obs. of  3 variables:
 $ name     : chr  "Alice Mutton" "Aniseed Syrup" "Boston Crab Meat" "Camembert
 $ category : chr  "Meat/Poultry" "Condiments" "Seafood" "Dairy Products" ...
 $ productid: int  1 2 3 4 5 6 7 8 9 10 ...
```

4 Assignment 4

4.1 a. What is good?

Let us begin by thinking of some characteristics that are valuable in a beta tester. A beta tester is valuable if he/she can test the entire game, thus he/she must be able to complete a game. Something else that is important is if the beta tester has experience with the genre of games that the to be tested game is in, so he/she can make good comparisons and give proper feedback. A lot of hours played, either overall or in beta testing, is also a plus, since the game needs to be extensively explored and tested. People who do not play a lot might not have the time or motivation to go through the entire game. A proper formula would be some weighted sum of those numbers.

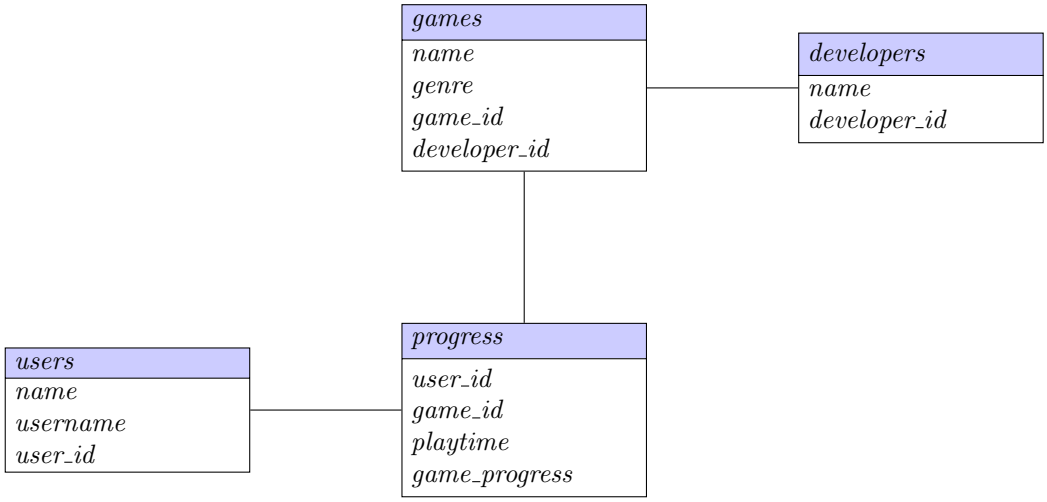
How good a certain player is, is less important to the overall equation. Since the game is (usually) targeted towards the general audience and not a niche group of expert players.

A (possible) problem with this approach is that the goodness score is dependent on the (genre of the) game. If we wish to eliminate this variable to obtain a constant goodness score, we would prefer a tester to have experience in many genres, since that makes the probability of relevant feedback higher

4.2 b. Business

For game developers: Who are the best qualified persons to test our game? The business question would be virtually the same for Pear, since the service lives or dies with its reputation of effectiveness. If Pear provides bad beta testers, then game developers will not sign up to the program.

4.3 c.


games
name
genre
game_id
developer_id
developers
name
developer_id
users
name
username
user_id
progress
user_id
game_id
playtime
game_progress