sentiment\_by\_category

Huaian Yu

June 17, 2019

### 0. Introduction

The data is from the Amazon’s apparel review data.(For more detailed information: [chuck@emadri.com](mailto:chuck@emadri.com))

I picked 1000 observations from the full dataset. Then extract adjetives from the “review\_body”. And judge the category of the product from the “product\_tittle”. The final dataset contains following variables:

-product\_id: ID of the product being reviewed.

-review\_id: The ID of reviewer.

-attributes: Describes the attributes of the variable “value”

-value: Words extraced from the customer review.

-count: How many time does a single word appears in a single review massage.

-tf: Term frequency.

-weight: How rare is the word.

-star\_rating: The rating for the product, given by the reviewer.

-item\_name: Categories of products.

-category: The category decided by me, for the covience of following analysis. Has 4 values: access(Accessories), top, bot and under(Underwear) (For more info: [andy@emadri.com](mailto:andy@emadri.com)) .

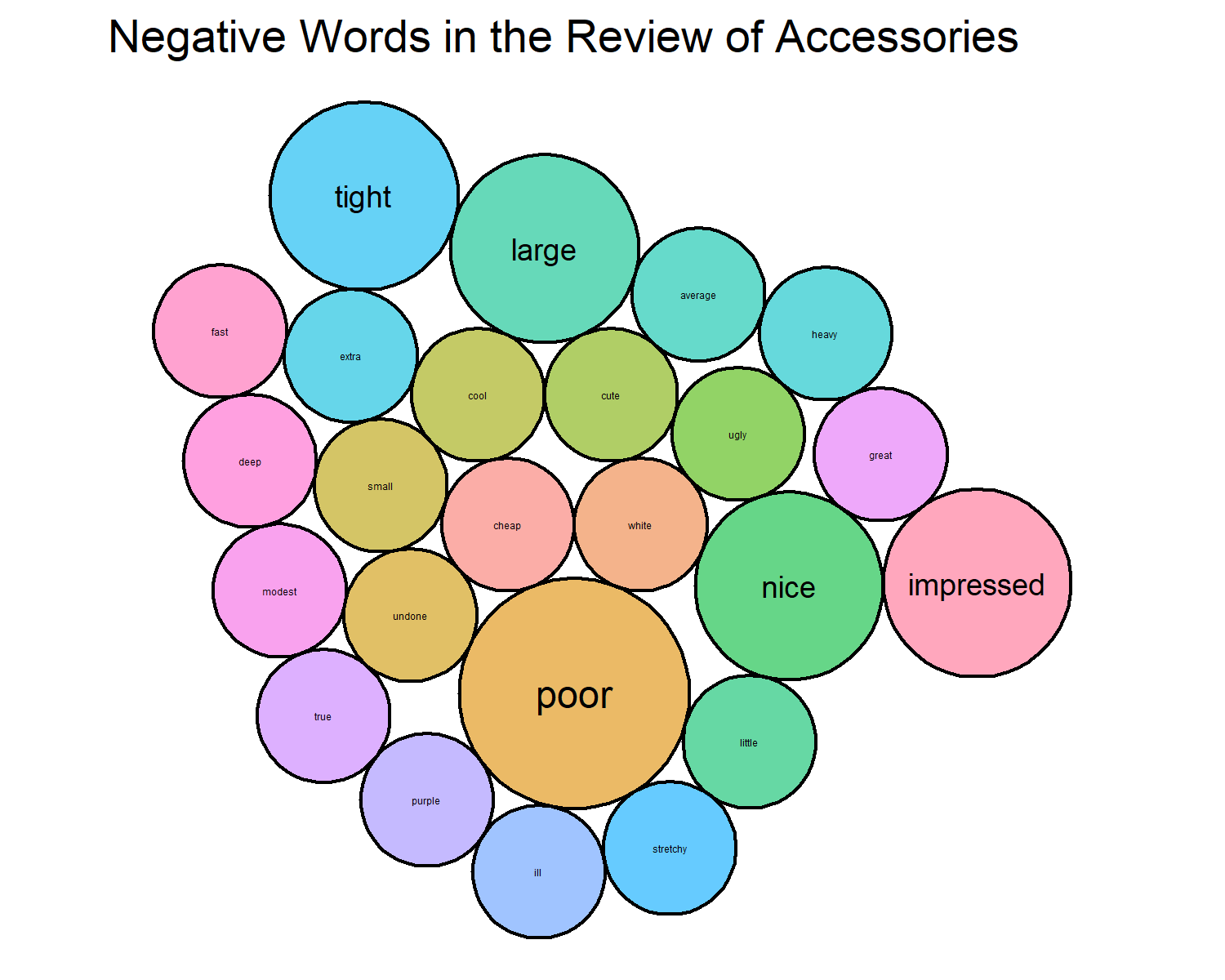
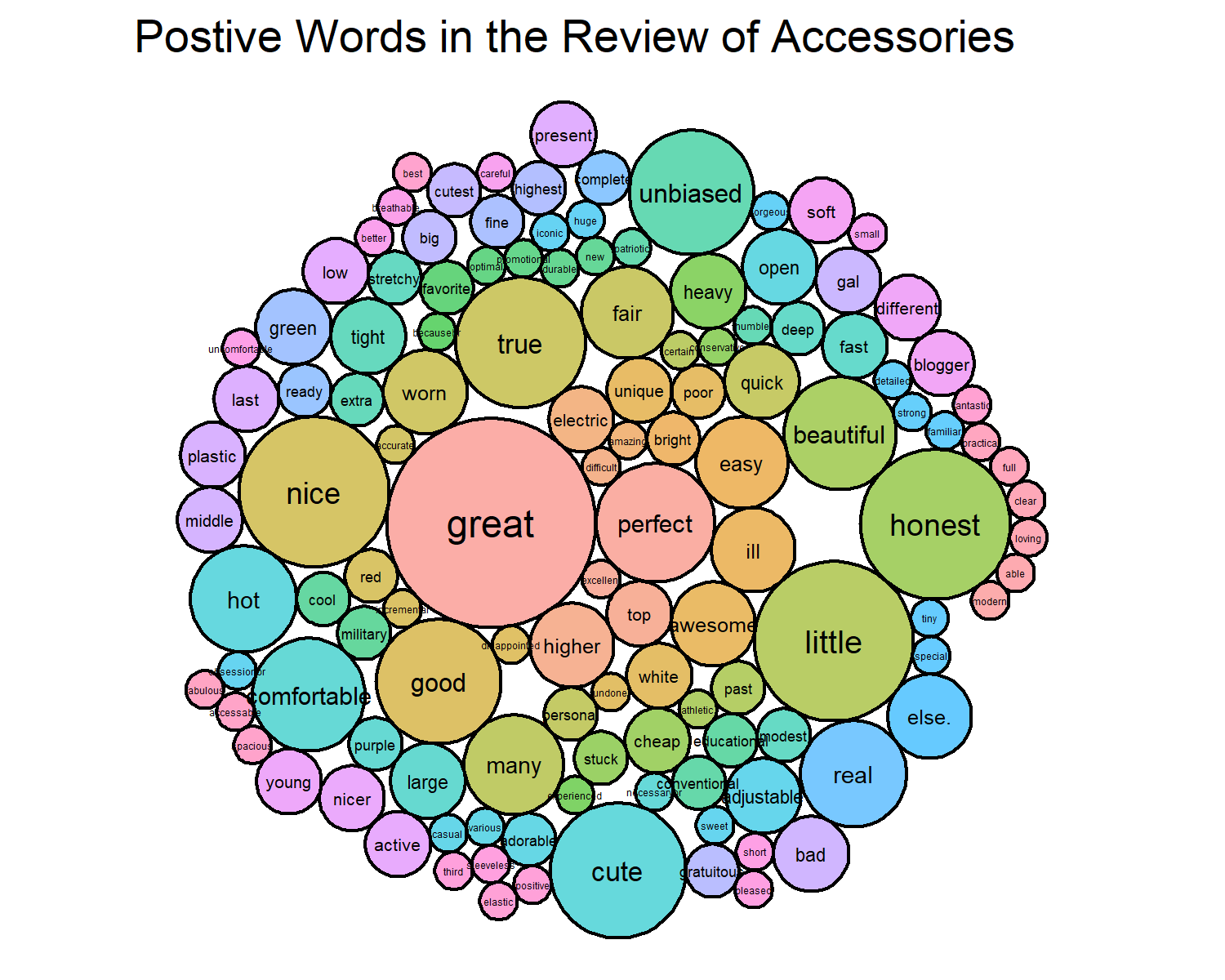
## product\_id review\_id attributes value count tf  
## 306 B013CTPBPE R3N8ZXJENRDD1I adj smalldisappointed 1 1.0000000  
## 307 B013CTPSJS R3VU0L98WUG5C2 adj nice 1 0.3333333  
## 308 B013CTPSJS R3VU0L98WUG5C2 comfort comfortable 1 0.3333333  
## 309 B013CTPSJS R3VU0L98WUG5C2 adj darker 1 0.3333333  
## 434 B013CUFO5K R3E1ZZ2VDGY74P adj great 1 1.0000000  
## 435 B013CV7H0O R1JMX9BBKAD6OB adj rough 1 0.0400000  
## 436 B013CV7H0O R1JMX9BBKAD6OB adj top 1 0.0400000  
## 437 B013CV7H0O R1JMX9BBKAD6OB adj concerned 1 0.0400000  
## 438 B013CV7H0O R1JMX9BBKAD6OB adj small 1 0.0400000  
## 439 B013CV7H0O R1JMX9BBKAD6OB adj long 1 0.0400000  
## weight star\_rating item\_name category  
## 306 6.82437367 1 cardigan top  
## 307 0.58515649 5 cardigan top  
## 308 0.64719058 5 cardigan top  
## 309 2.04374216 5 cardigan top  
## 434 1.31498533 5 cap access  
## 435 0.21752317 4 shirts top  
## 436 0.10599946 4 shirts top  
## 437 0.21752317 4 shirts top  
## 438 0.08147528 4 shirts top  
## 439 0.11055723 4 shirts top

The purpose of this first step analysis, is to find a set of keywords that are associated with possitive review. Then, to see if these key words could show some insight about how could a customer satisfy with his/her purchase.

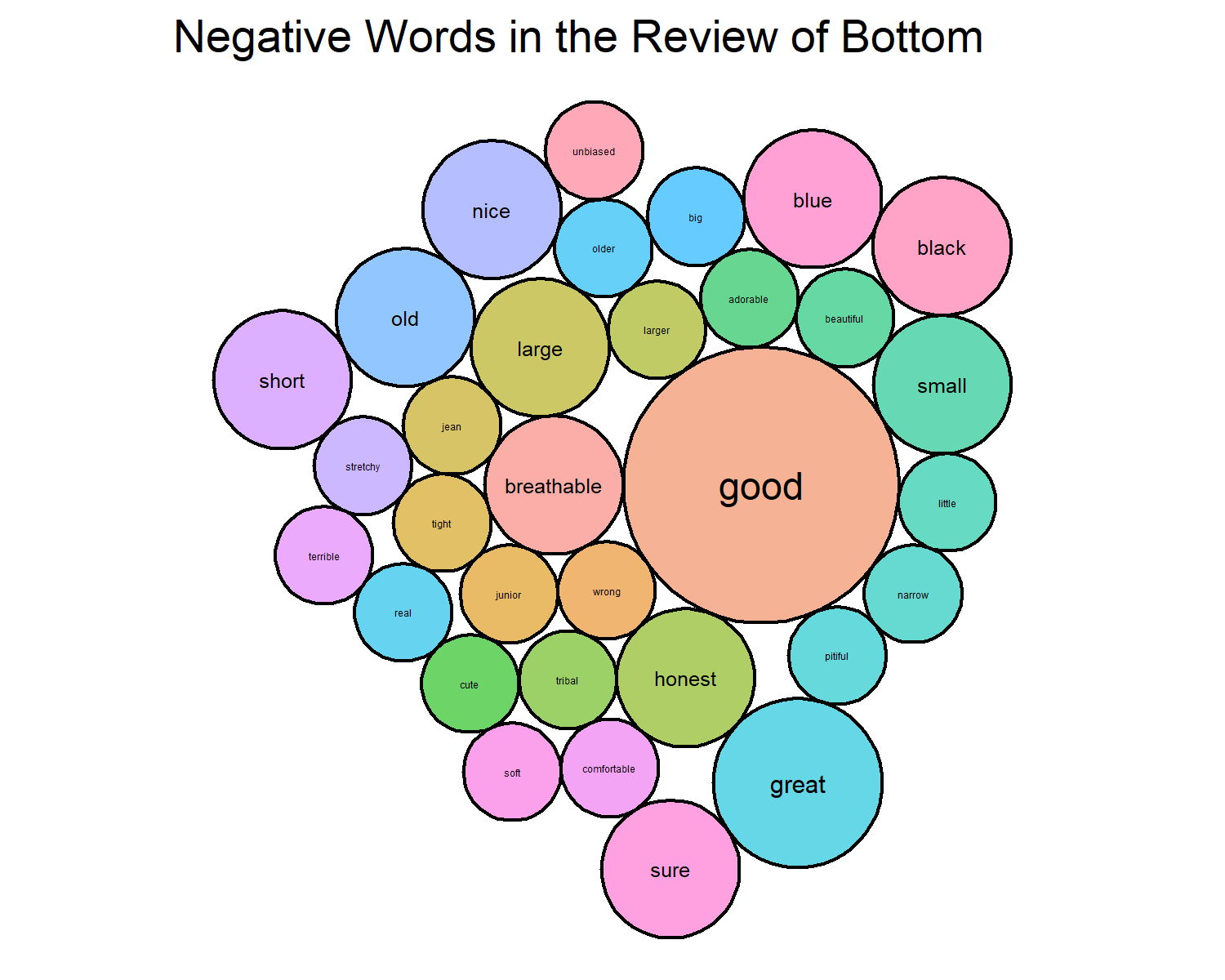
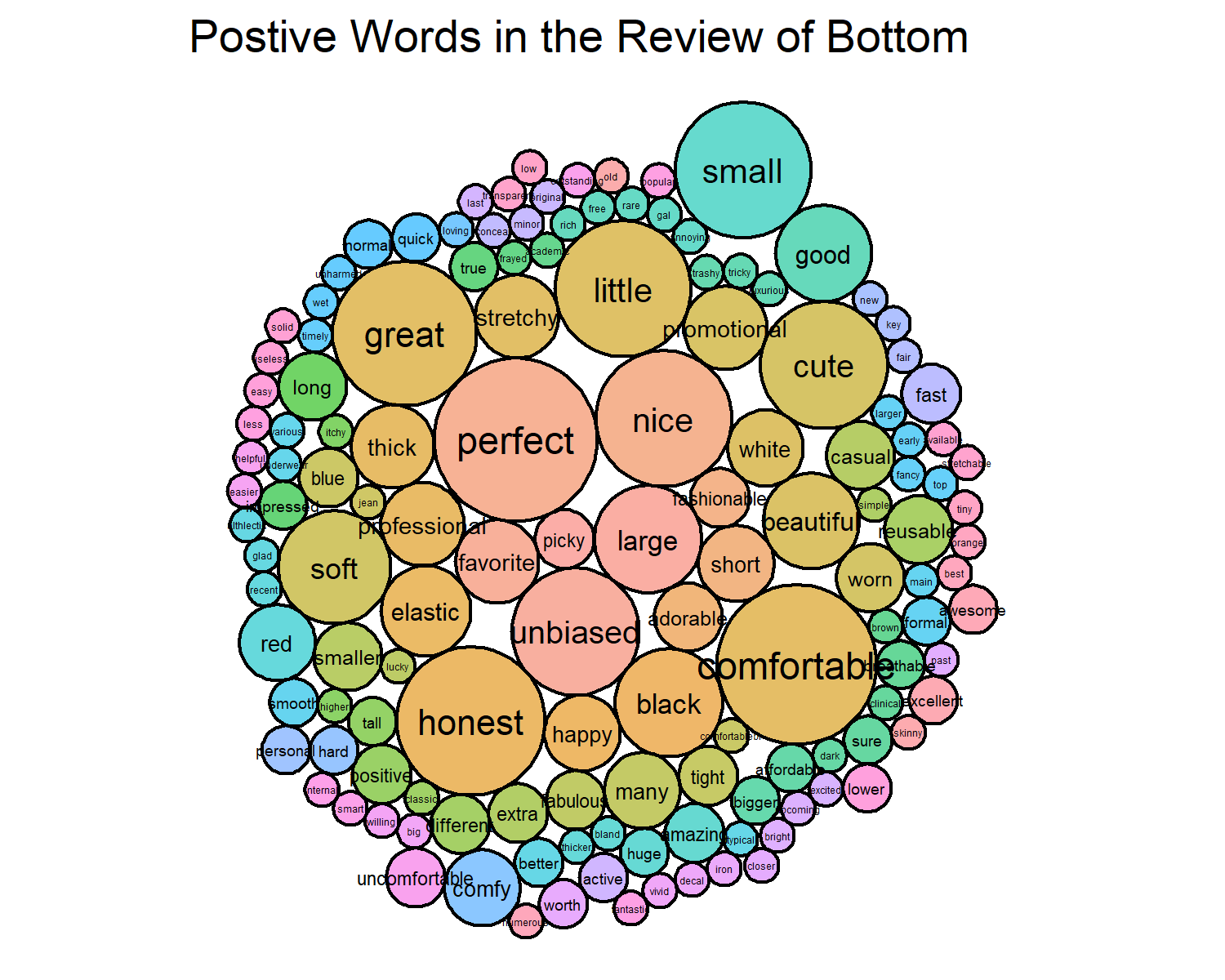
Here, I divide extracted into two groups: possitive and negative. Posstive words are from reviews rated with more than 3 stars. Negative words are from review rated with less than 3 stars.

In the following plots, words with larger circle means they are more frequently appeared in the review.

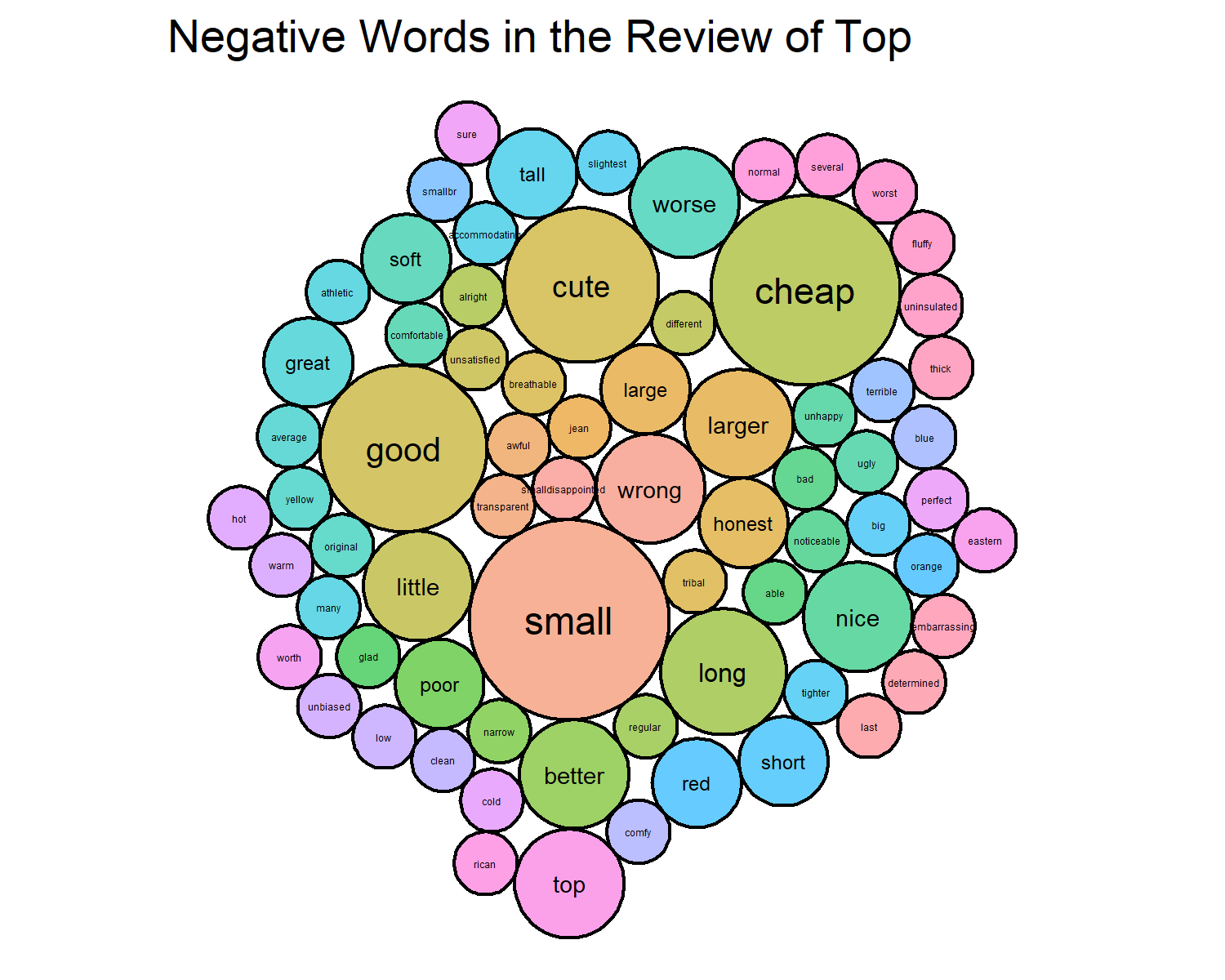
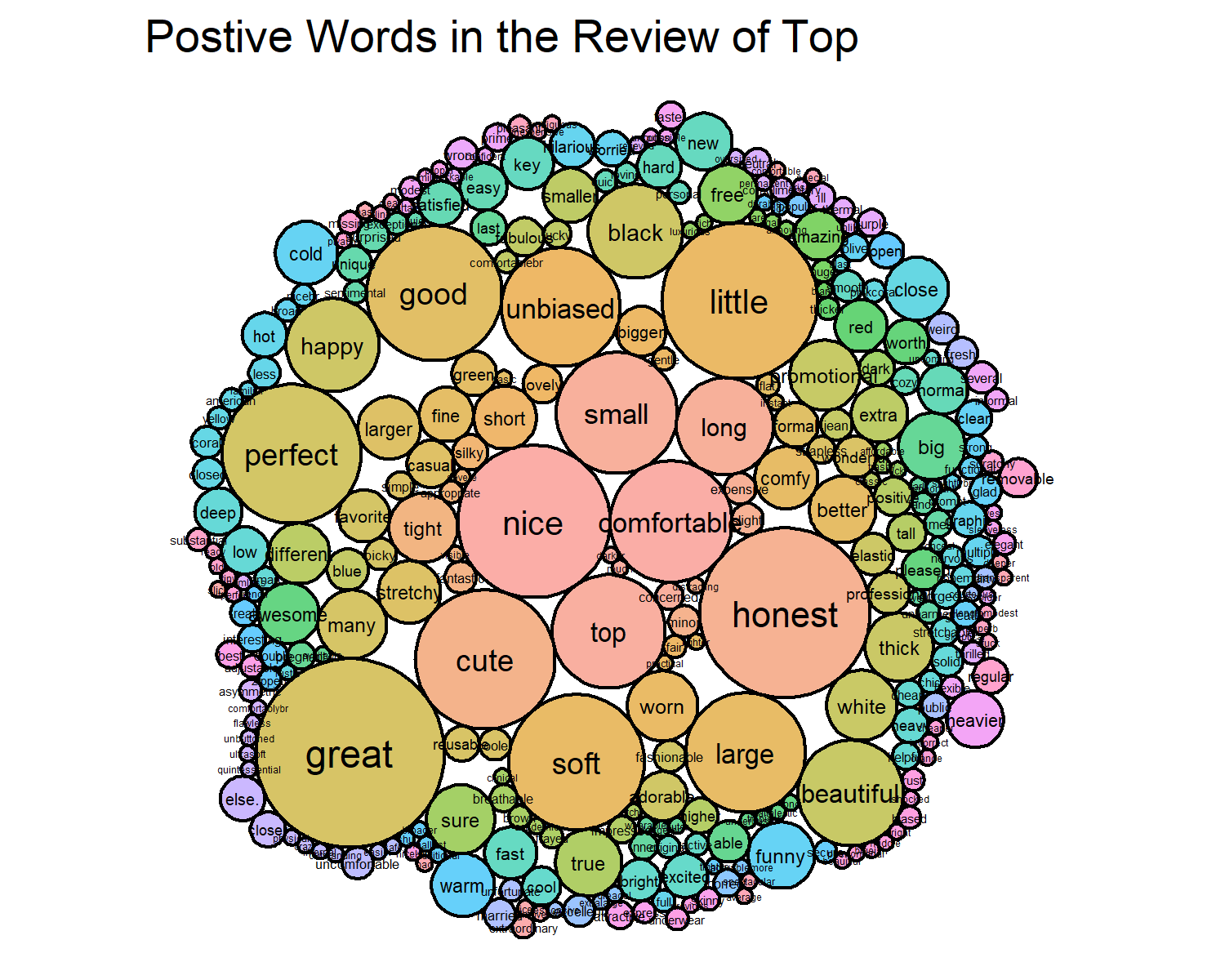
### 1. Accessories



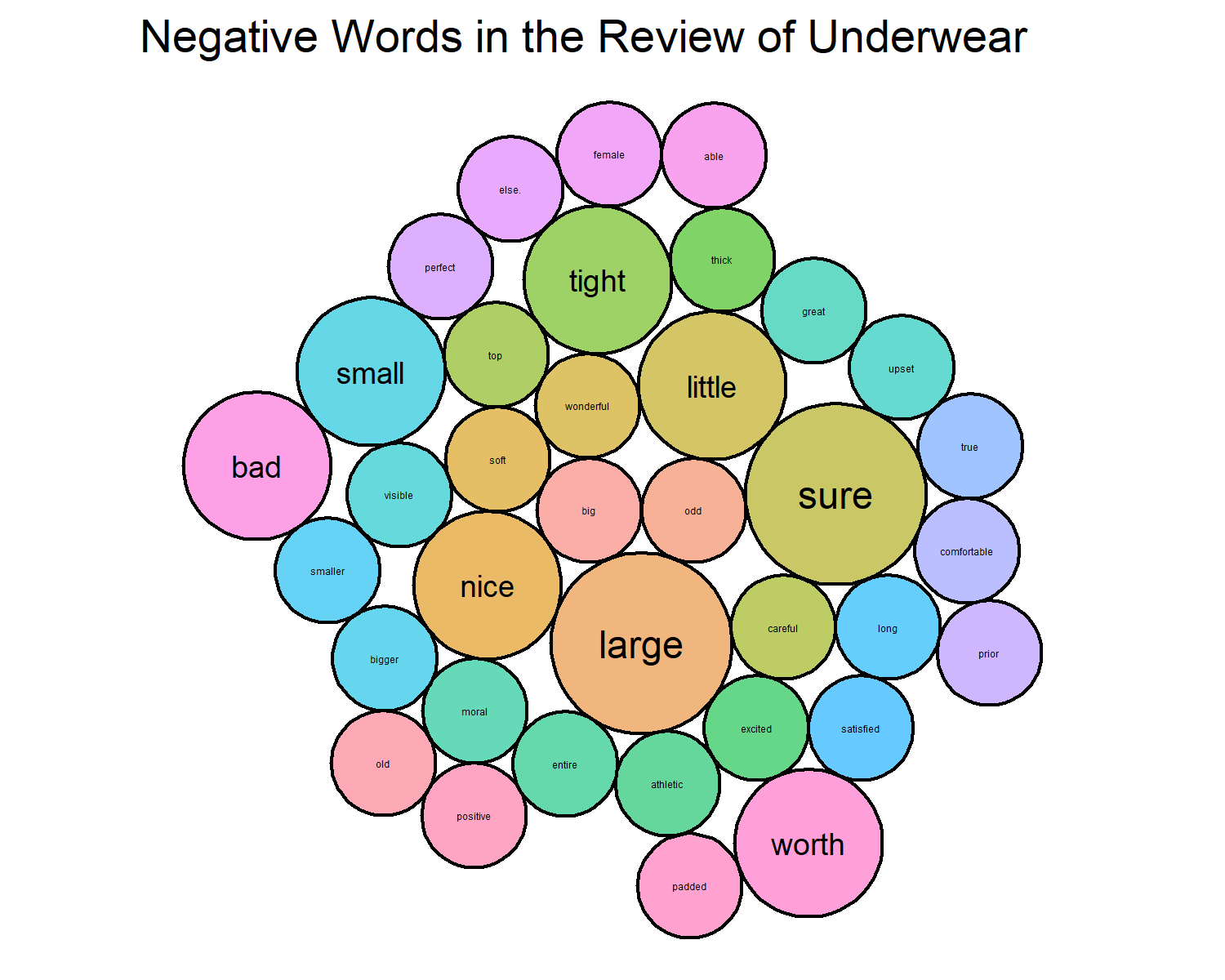
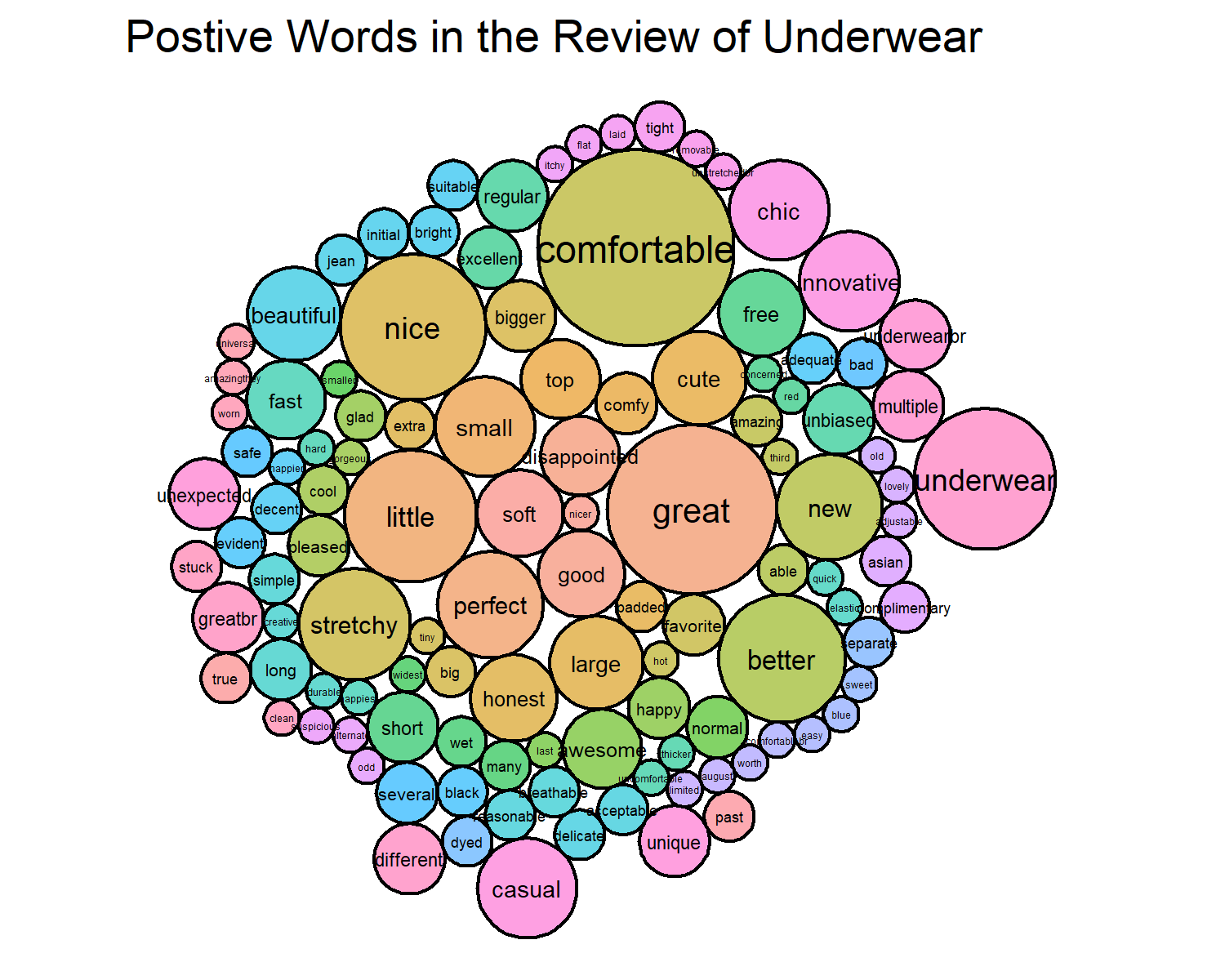
### 2. Bottom



### 3. Top



### 4. Underwear



# 5. Summary

In general, there are 3 perspectives that affact the review of a product:

1. How does the product look like:

Example words: unique, beautiful, perfect, great, excellent, good, hot, nice, cute, awesome, fashionable, chic

1. How’s it feel to wear the product on:

Example words: not heavy, soft, comfortable, adjustable, not too tight, not too large, stretchy, comfy, breathable

1. How’s the service offered by the seller:

Example words: unbiased(description on the size/color of the product), honest, real, happy, promotional