

# 628888

*Liqiang Pu*

data clean:

```
library(recommenderlab)
```

```
## Warning: package 'recommenderlab' was built under R version 3.2.3
```

```
## Loading required package: Matrix
```

```
## Loading required package: registry
```

```
## Loading required package: arules
```

```
## Warning: package 'arules' was built under R version 3.2.3
```

```
##  
## Attaching package: 'arules'
```

```
## The following objects are masked from 'package:base':  
##  
##      %in%, abbreviate, write
```

```
## Loading required package: proxy
```

```
##  
## Attaching package: 'proxy'
```

```
## The following object is masked from 'package:Matrix':  
##  
##      as.matrix
```

```
## The following objects are masked from 'package:stats':  
##  
##      as.dist, dist
```

```
## The following object is masked from 'package:base':  
##  
##      as.matrix
```

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:arules':  
##  
## intersect, setdiff, setequal, union
```

```
## The following objects are masked from 'package:stats':  
##  
## filter, lag
```

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

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 3.2.3
```

```
##  
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:Matrix':  
##  
## expand
```

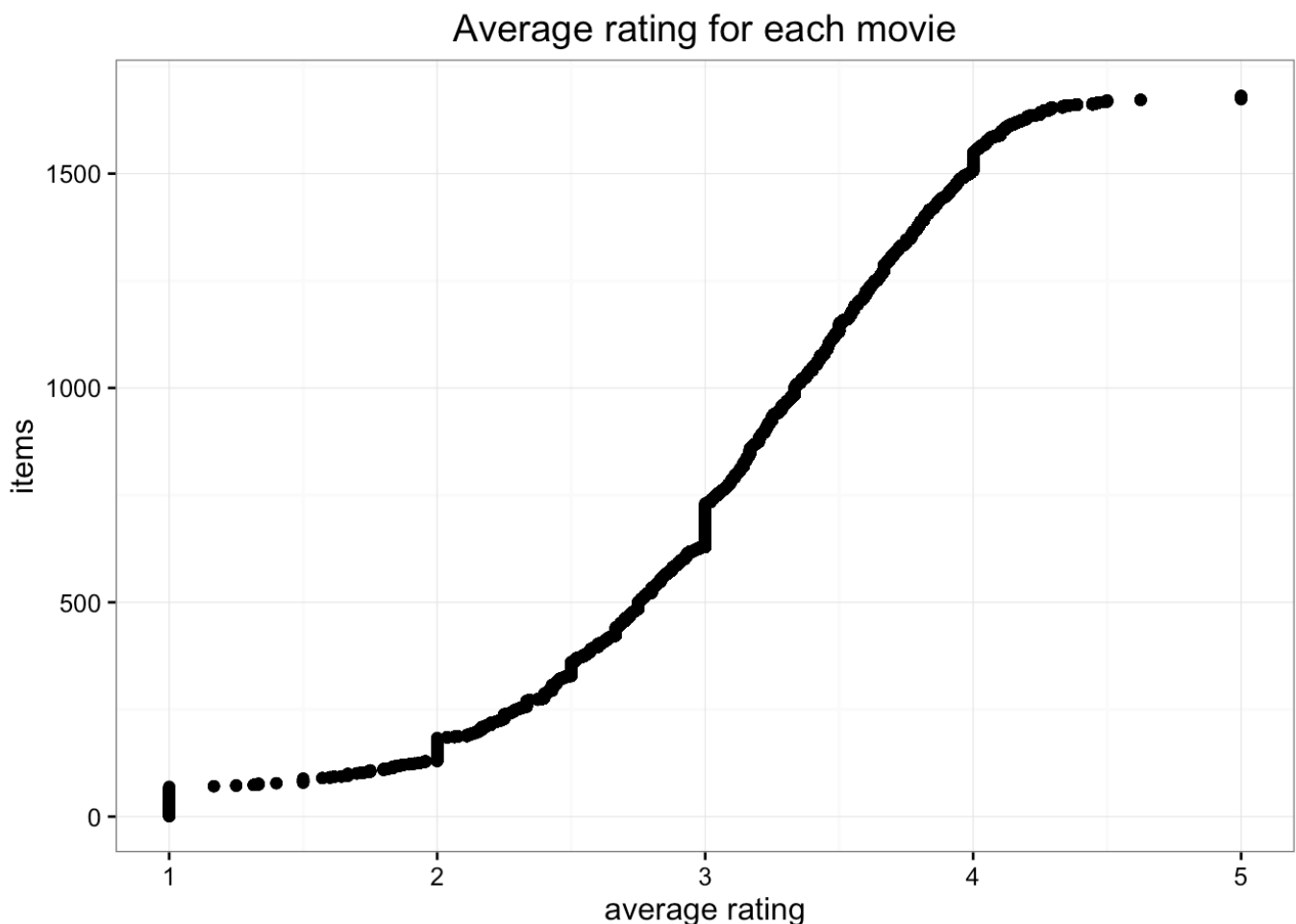
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.3
```

```
#100K data source #http://files.grouplens.org/datasets/movielens/ml-100k  
movie<-read.table("http://files.grouplens.org/datasets/movielens/ml-100k/u.dat  
a")  
colnames(movie)<-c("users","items","rating","timestamp")  
users<-read.table("http://files.grouplens.org/datasets/movielens/ml-100k/u.use  
r",sep="|")  
colnames(users)<-c("userid","age","gender","occupation","zipcode")  
item<-read.delim("http://files.grouplens.org/datasets/movielens/ml-100k/u.ite  
m",sep="|",header=F)  
item<-item[,c(-1,-3,-4,-5)]  
colnames(item)<-c("MovieName","unknown", "Action","Adventure","Animation","Chil  
dren", "Comedy", "Crime", "Documentary","Drama","Fantasy","Film-Noir","Horro  
r","Musical","Mystery","Romance","Sci-Fi","Thriller","War","Western")
```

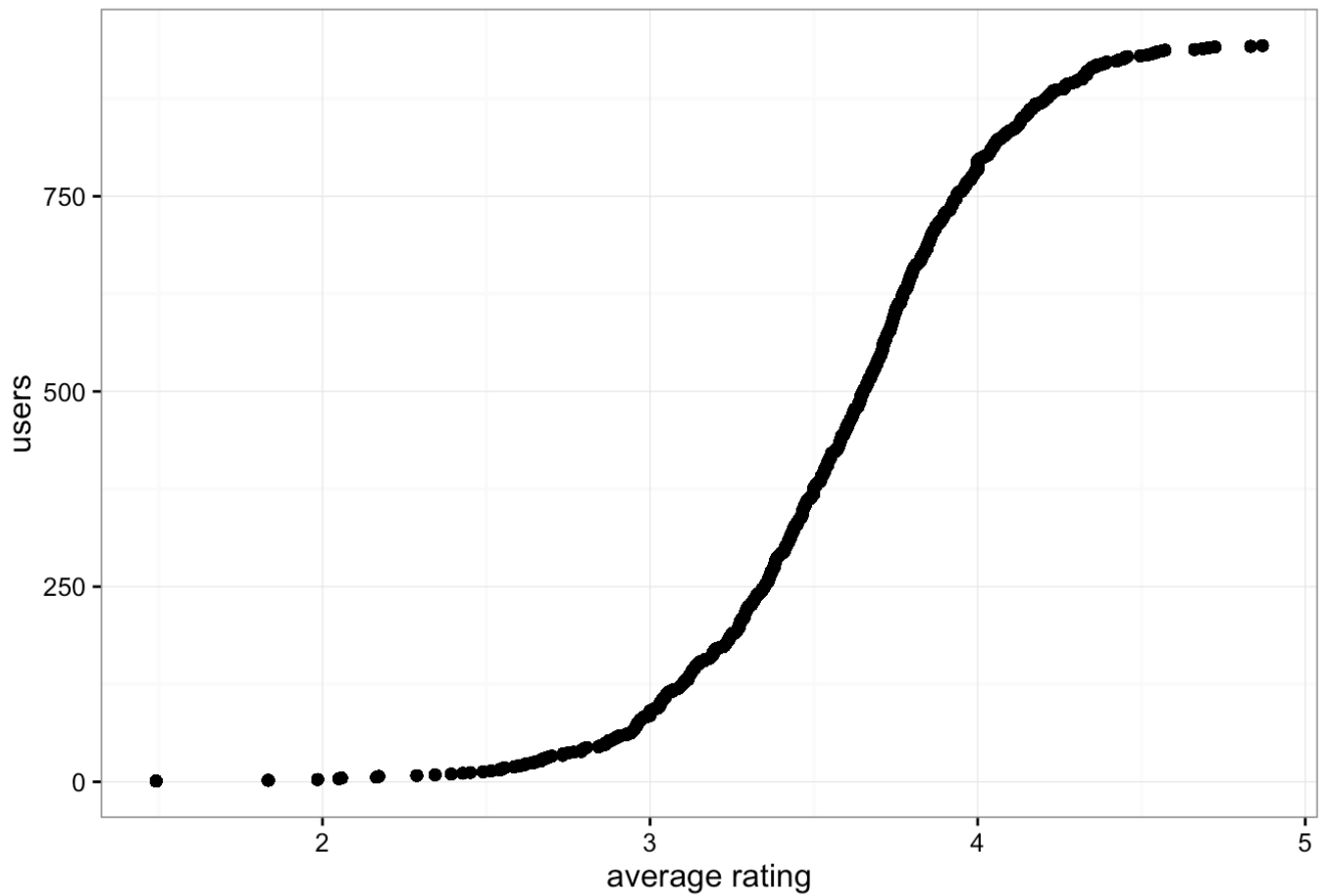
## Data visualization:

```
movie_item<-movie %>%
  group_by(items) %>%
  mutate(averating_item=mean(rating)) %>%
  ungroup %>%
  mutate(items = reorder(items, averating_item))
ggplot(movie_item,aes(x=averating_item,y=as.numeric(items)))+geom_point()+ylab(
  ("items")+xlab("average rating")+ggtitle("Average rating for each movie")+ theme
  _bw()
```

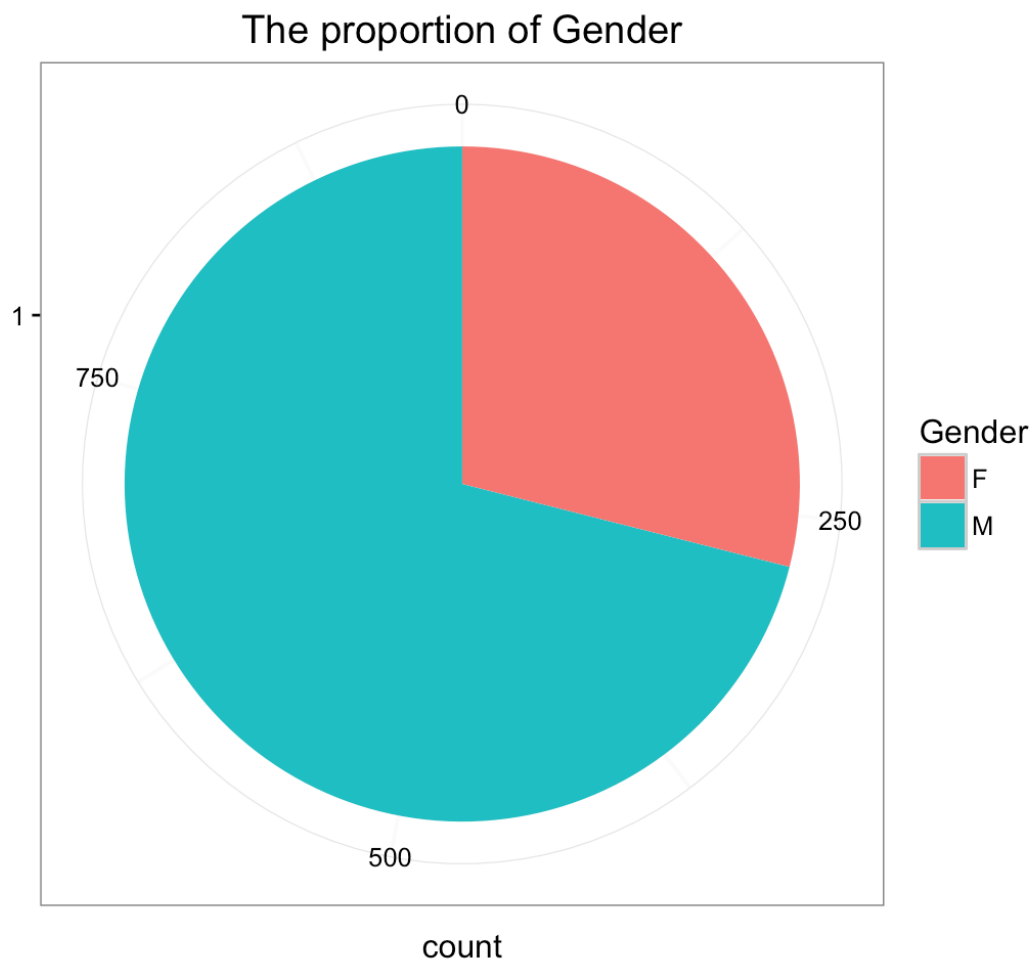


```
movie_user<-movie %>%
  group_by(users) %>%
  mutate(averating_users=mean(rating)) %>%
  ungroup %>%
  mutate(users = reorder(users, averating_users))
ggplot(movie_user,aes(x=averating_users,y=as.numeric(users)))+geom_point()+ylab(
  ("users")+xlab("average rating")+ggtitle("Average rating for each user")+ theme
  _bw()
```

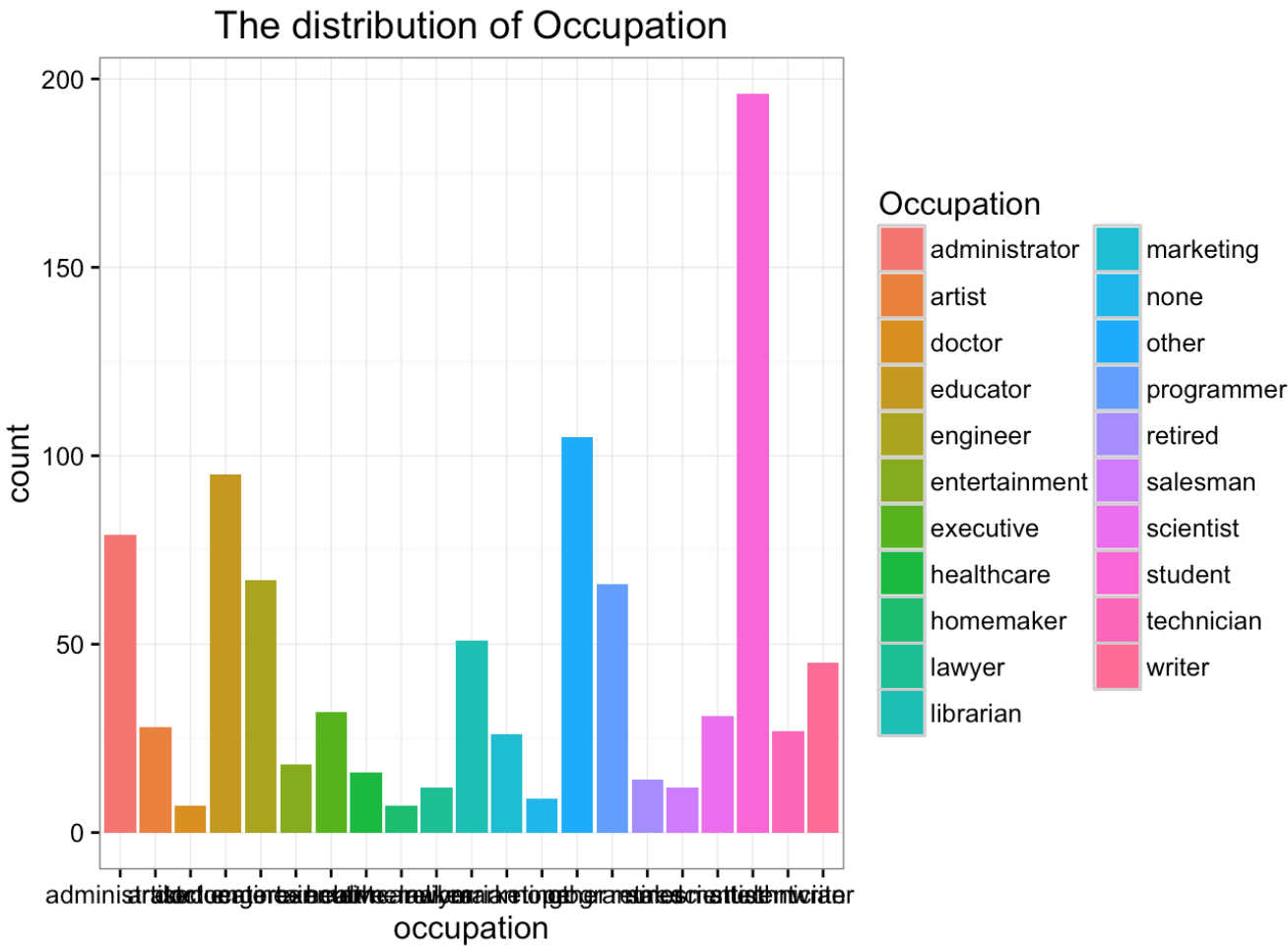
Average rating for each user



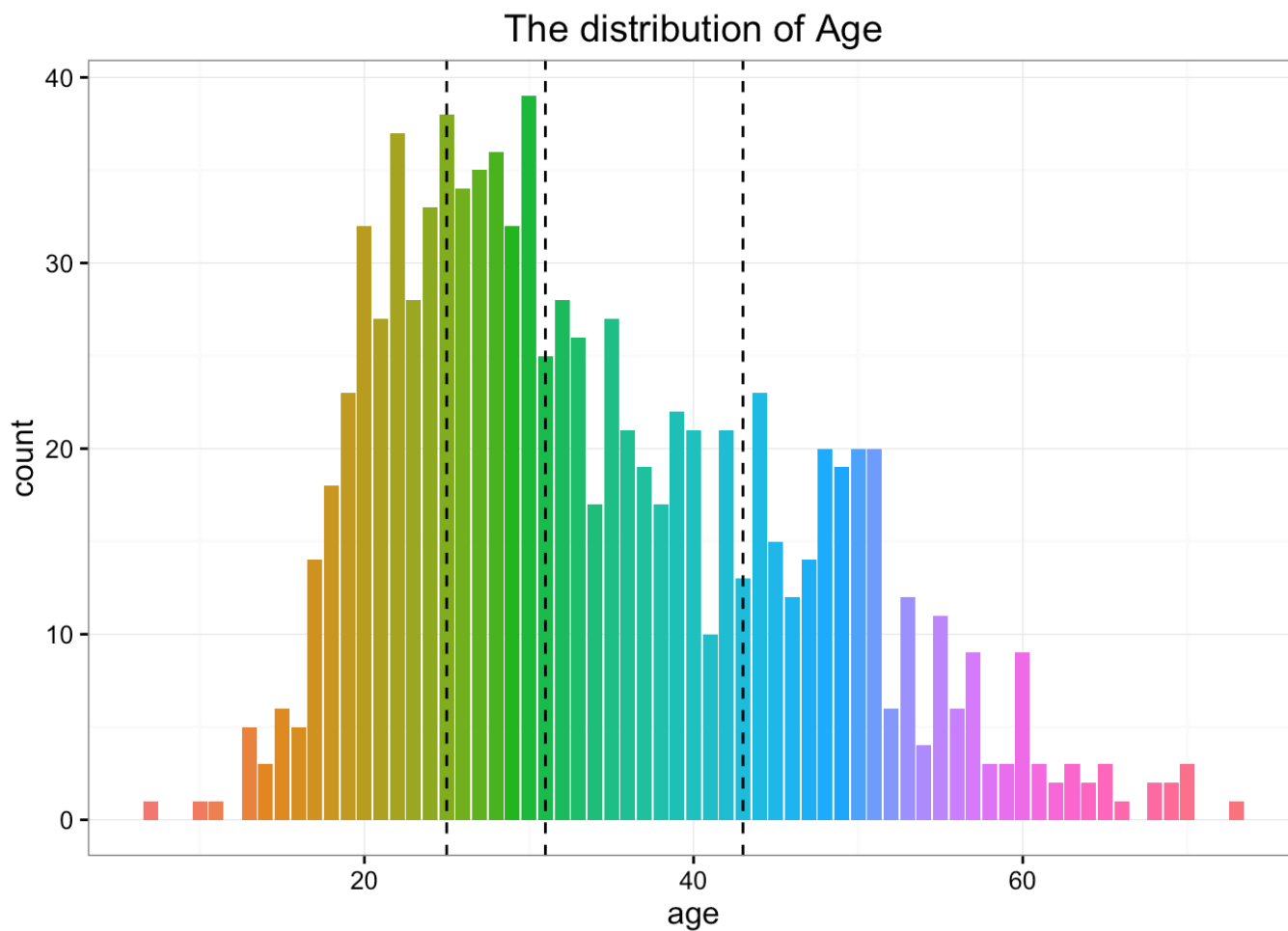
```
ggplot(users, aes(x =factor(1), fill = factor(gender)))+geom_bar(width = 1)+ coord_polar(theta = "y")+scale_fill_discrete(guide=guide_legend(title = "Gender"))+ggtitle("The proportion of Gender")+xlab("")+ theme_bw()
```



```
ggplot(users, aes(x =occupation, fill =factor(occupation)))+geom_bar()+scale_fill_discrete(guide=guide_legend(title = "Occupation"))+ggtitle("The distribution of Occupation")+ theme_bw()
```



```
agelevel<-quantile(users$age, probs = seq(0, 1, 0.25))
ggplot(users, aes(x =age, fill =factor(age)))+geom_bar(show.legend = F)+ggtitle
("The distribution of Age")+geom_vline(xintercept=agelevel[c(2,3,4)],linetype =
2)+ theme_bw()
```



```

moviematrix<-as(item[,-1],"matrix")
moviematrix<-as(moviematrix, "binaryRatingMatrix")
rec=Recommender(moviematrix,method="POPULAR")
as(getModel(rec)$topN,"list")

```

```

## [[1]]
##  [1] "Drama"      "Comedy"     "Action"     "Thriller"   "Romance"
##  [6] "Adventure"  "Children"   "Crime"      "Sci-Fi"     "Horror"
## [11] "War"        "Mystery"    "Musical"    "Documentary" "Animation"
## [16] "Western"    "Film-Noir" "Fantasy"    "unknown"

```

```

# Most popular movie types are Drama , Comedy, Thriller & Action

```