BIRMINGHAM CITY
University

CMP5352 Report - TITLE NEEDED

DRAFT VERSION

Lewis Higgins - Student ID 22133848

Word count: XXXX

Abstract

text
    text
    text

Contents

Introduction

Text text text

    Text text text

Motivation and objectives

```
ggplot(diamonds, aes(x = carat, fill = color)) +
  geom_histogram(bins = 30)
```
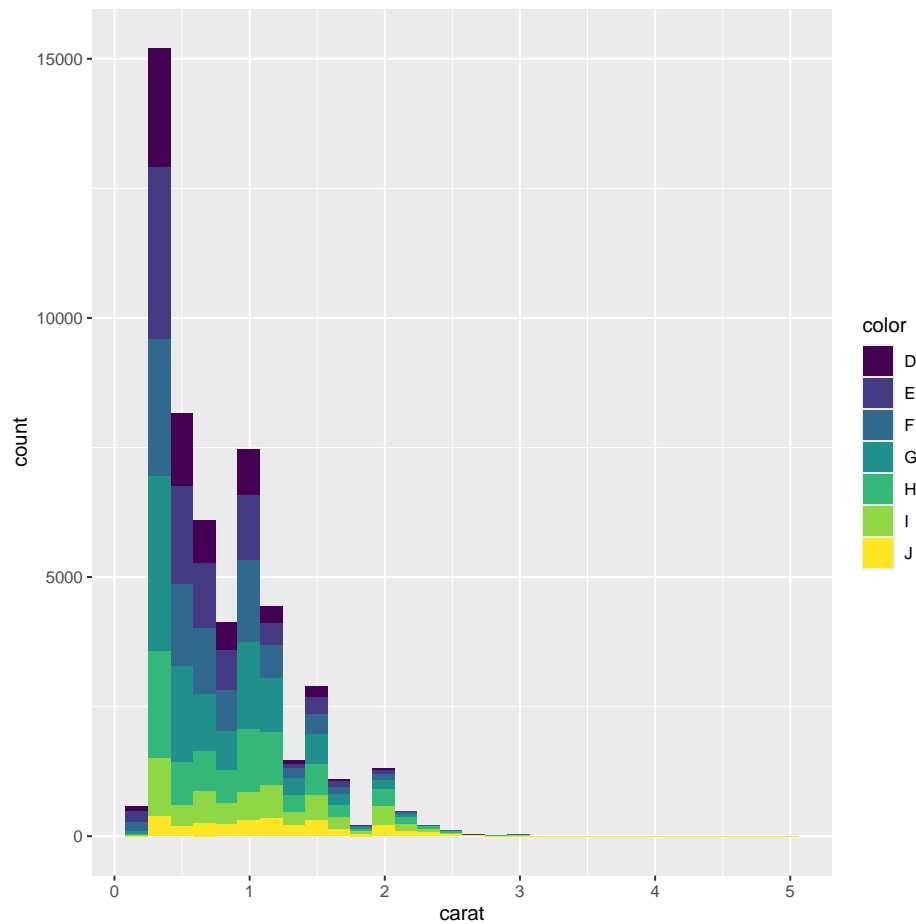

Figure 1: plotting example

As we can see in figure 1

```
ggplot(diamonds, aes(x = carat, y = price)) +
geom_point(aes(color = cut))
```
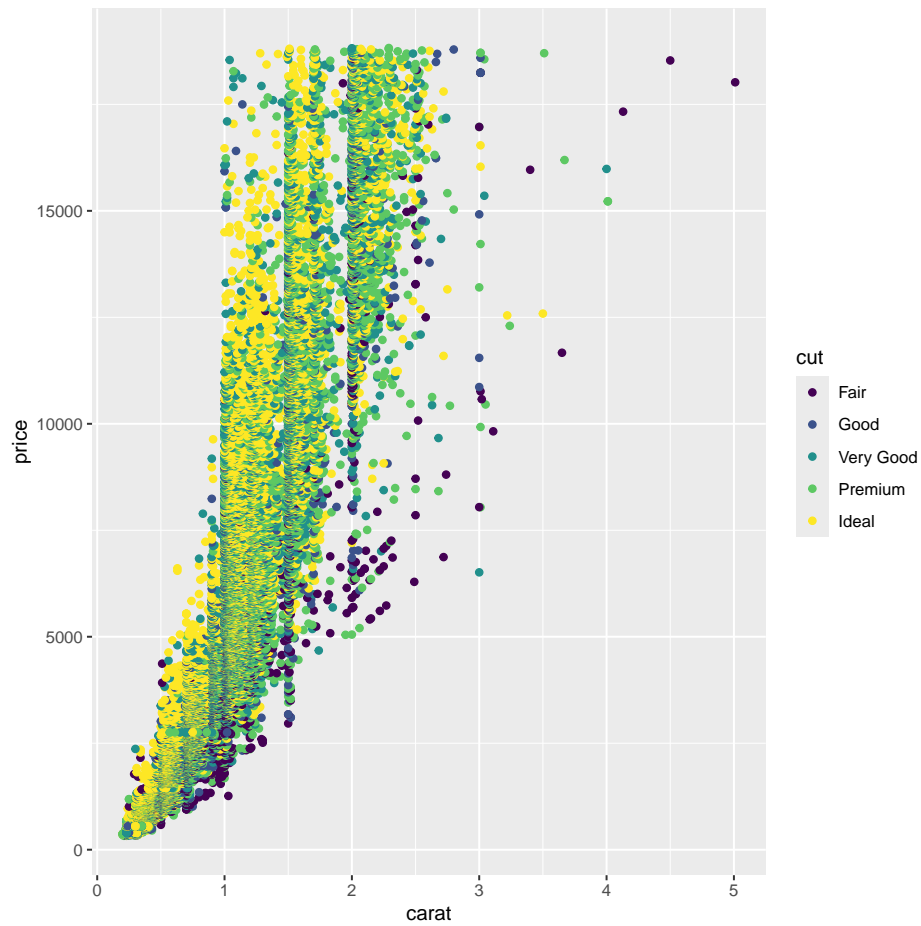

Figure 2: plotting example 2

Also from figure 2

Experimental results

```r
summary(dataDf)
```

```
##                                                App               Category
##   ROBLOX                                        :    9    FAMILY       :1972
##   CBS Sports App - Scores, News, Stats & Watch Live:    8    GAME         :1144
##   8 Ball Pool                                   :    7    TOOLS        : 843
##   Candy Crush Saga                              :    7    MEDICAL      : 463
##   Duolingo: Learn Languages Free                :    7    BUSINESS     : 460
##   ESPN                                          :    7    PRODUCTIVITY: 424
##   (Other)                                       :10796    (Other)      :5535
##      Rating        Reviews                 Size              Installs
##   Min.   : 1.0   0      : 596   Varies with device:1695   1,000,000+ :1579
##   1st Qu.: 4.0   1      : 272   11M               : 198   10,000,000+:1252
##   Median : 4.3   2      : 214   12M               : 196   100,000+   :1169
##   Mean   : 4.2   3      : 175   14M               : 194   10,000+    :1054
##   3rd Qu.: 4.5   4      : 137   13M               : 191   1,000+     : 907
##   Max.   :19.0   5      : 108   15M               : 184   5,000,000+ : 752
##   NA's   :1474   (Other):9339   (Other)           :8183   (Other)    :4128
##     Type         Price              Content.Rating          Genres
##   0   :    1   0      :10040                  :   1   Tools        : 842
##   Free:10039   $0.99  :  148   Adults only 18+:   3   Entertainment: 623
##   NaN :    1   $2.99  :  129   Everyone       :8714   Education    : 549
##   Paid:  800   $1.99  :   73   Everyone 10+   : 414   Medical      : 463
##                $4.99  :   72   Mature 17+     : 499   Business     : 460
##                $3.99  :   63   Teen           :1208   Productivity : 424
##                (Other):  316   Unrated        :   2   (Other)      :7480
##         Last.Updated          Current.Ver               Android.Ver
##   August 3, 2018: 326   Varies with device:1459   4.1 and up        :2451
##   August 2, 2018: 304   1.0               : 809   4.0.3 and up      :1501
##   July 31, 2018 : 294   1.1               : 264   4.0 and up        :1375
##   August 1, 2018: 285   1.2               : 178   Varies with device:1362
##   July 30, 2018 : 211   2.0               : 151   4.4 and up        : 980
##   July 25, 2018 : 164   1.3               : 145   2.3 and up        : 652
##   (Other)       :9257   (Other)           :7835   (Other)           :2520
```

```r
# Replace all instances of 'NA' as a string or a blank string "" to NA.
# The strings 'NA' or " " and NA are two different things to R, and only the
# non-string NA is detected by functions like is.na().
dataDf[(dataDf == 'NA' | dataDf == "" | dataDf == "NaN")] <- NA

# Output a summary of the data.
summary(dataDf)
```

```
##                                                App               Category
```

```
##   ROBLOX                                    :    9   FAMILY      :1972
##   CBS Sports App - Scores, News, Stats & Watch Live:    8   GAME        :1144
##   8 Ball Pool                               :    7   TOOLS       : 843
##   Candy Crush Saga                          :    7   MEDICAL     : 463
##   Duolingo: Learn Languages Free            :    7   BUSINESS    : 460
##   ESPN                                      :    7   PRODUCTIVITY: 424
##   (Other)                                   :10796   (Other)     :5535
##      Rating        Reviews                 Size            Installs
##   Min.   : 1.0   0      : 596   Varies with device:1695   1,000,000+ :1579
##   1st Qu.: 4.0   1      : 272   11M               : 198   10,000,000+:1252
##   Median : 4.3   2      : 214   12M               : 196   100,000+   :1169
##   Mean   : 4.2   3      : 175   14M               : 194   10,000+    :1054
##   3rd Qu.: 4.5   4      : 137   13M               : 191   1,000+     : 907
##   Max.   :19.0   5      : 108   15M               : 184   5,000,000+ : 752
##   NA's   :1474   (Other):9339   (Other)           :8183   (Other)    :4128
##      Type          Price           Content.Rating         Genres
##   0   :    1   0      :10040   Everyone     :8714   Tools        : 842
##   Free:10039   $0.99  :  148   Teen         :1208   Entertainment: 623
##   NaN :    0   $2.99  :  129   Mature 17+   : 499   Education    : 549
##   Paid:  800   $1.99  :   73   Everyone 10+ : 414   Medical      : 463
##   NA's:    1   $4.99  :   72   Adults only 18+:   3   Business     : 460
##                $3.99  :   63   (Other)      :   2   Productivity : 424
##                (Other):  316   NA's         :   1   (Other)      :7480
##      Last.Updated          Current.Ver           Android.Ver
##   August 3, 2018: 326   Varies with device:1459   4.1 and up    :2451
##   August 2, 2018: 304   1.0               : 809   4.0.3 and up  :1501
##   July 31, 2018 : 294   1.1               : 264   4.0 and up    :1375
##   August 1, 2018: 285   1.2               : 178   Varies with device:1362
##   July 30, 2018 : 211   2.0               : 151   4.4 and up    : 980
##   July 25, 2018 : 164   (Other)           :7972   (Other)       :3169
##   (Other)       :9257   NA's              :   8   NA's          :   3
```

```r
# Identify rows containing NA.
naRows <- dataDf[rowSums(is.na(dataDf)) > 0,]
# 1469 with no ratings, 8 with no current version, 4 with neither
nrow(naRows)
```

```
## [1] 1481
```

Summary & conclusion

aaa