Report about E-colored Diamonds

Hanna Tillmanns (WIdO)

Stand: 14.06.2019

1 Defining parameters “outside” the script

To be able to render the scripts with different parameters and not having to start it by hand, I use the “params”-value of the YAML.

The scripts are then startet with rmarkdown::render.

2 Define all the general things and import the data

First things first. At the beginning I load all libraries, define the general parameters and import all of the data. In my case importing the data takes some time, therefore it is nice not having to repeat this for every single child-script, especially those that run several times.

```
# define parames
col<- params$col


#load librarys

library(openxlsx)
#library(reshape2)
#library (RODBC)
#library(lmtest)
#library(ggplot2)
library(foreach)
library(tidyverse)
#library(treemap)
#library(gridExtra)
#library(grid)
library(knitr)
#library(pander)
#library(kableExtra)
#library(captioner)
#library(here)
#library(RGraphics)
#library(treemapify)


#Creating the Data for the Report

data <- diamonds %>%
  filter(color==col)
```

3 Creating the Report

The report can be divided into different types of sections:

- all those sections that are within the parent script
- sections that are in child-scripts
- without a loop and without a condition
- with a condition that has to be true to render the file
- within a loop, which creates a section for every defined specificaton.

4 Single child-script

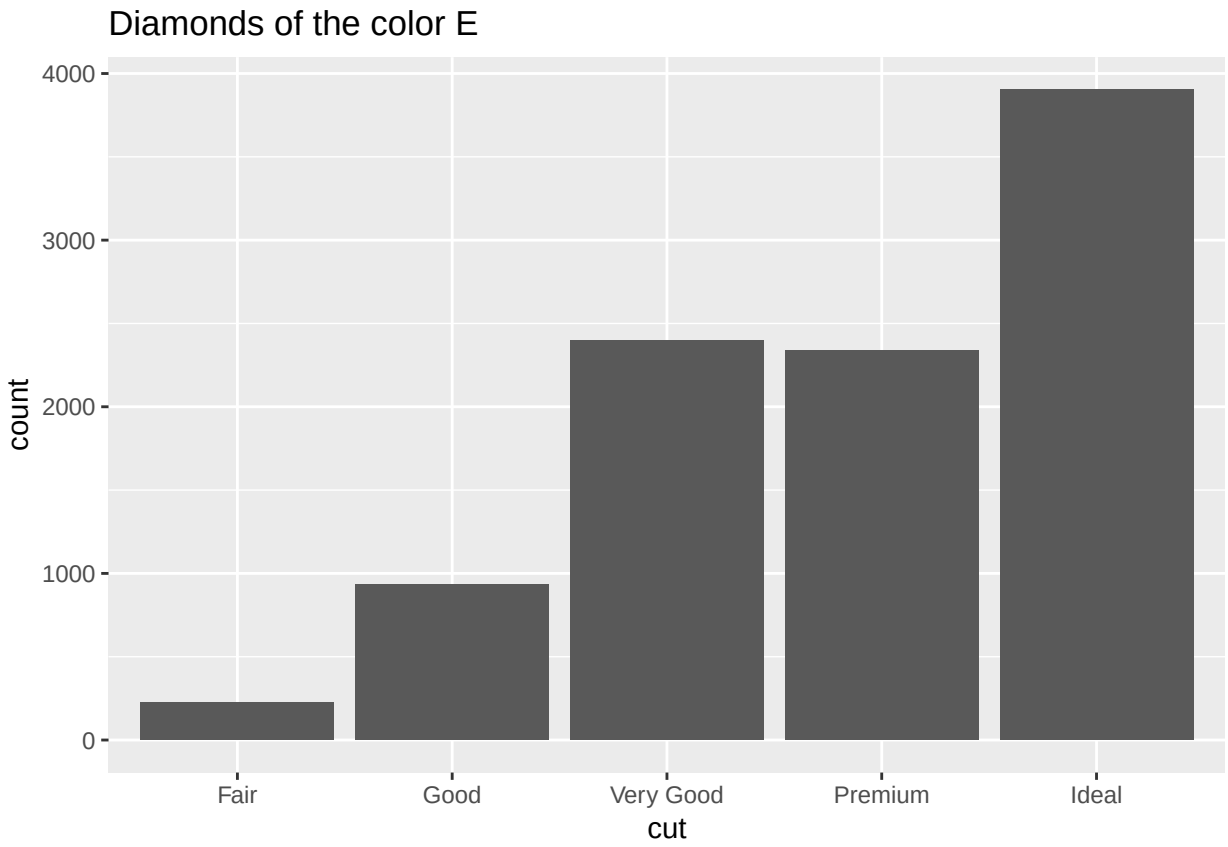
```
# calling a single child with

#```{r single_child,child = 'Single_Child.rmd'}
#```
```

4.1 Start Child-Script

The section gets all the parameters and data from the parent script.

Myselfe use it mainly for organising my script.

```
  ggplot(data=data) +
  geom_bar(aes(x=cut)) +
  labs(title=paste0("Diamonds of the color ",col))
```



5 Conditional child-script

The script will only be evaluated if the condition is true. This is a nice way to cope for example with missing data.

```
#creating the test

test <- ifelse(data %>%
                  filter(clarity=="SI2") %>% nrow()>0,TRUE,FALSE)
```

```
#calling the conditional child-script with:

#```{r conditional_print_script, child='cond_child.Rmd', eval = test}
#```
```

5.1 Start conditional child-script

The average price of diamonds with the clarity “SI2” is 4173.8260362.

6 Looping over child-scripts

As the names of chunks may not be repeated, the chunks should be nameless.

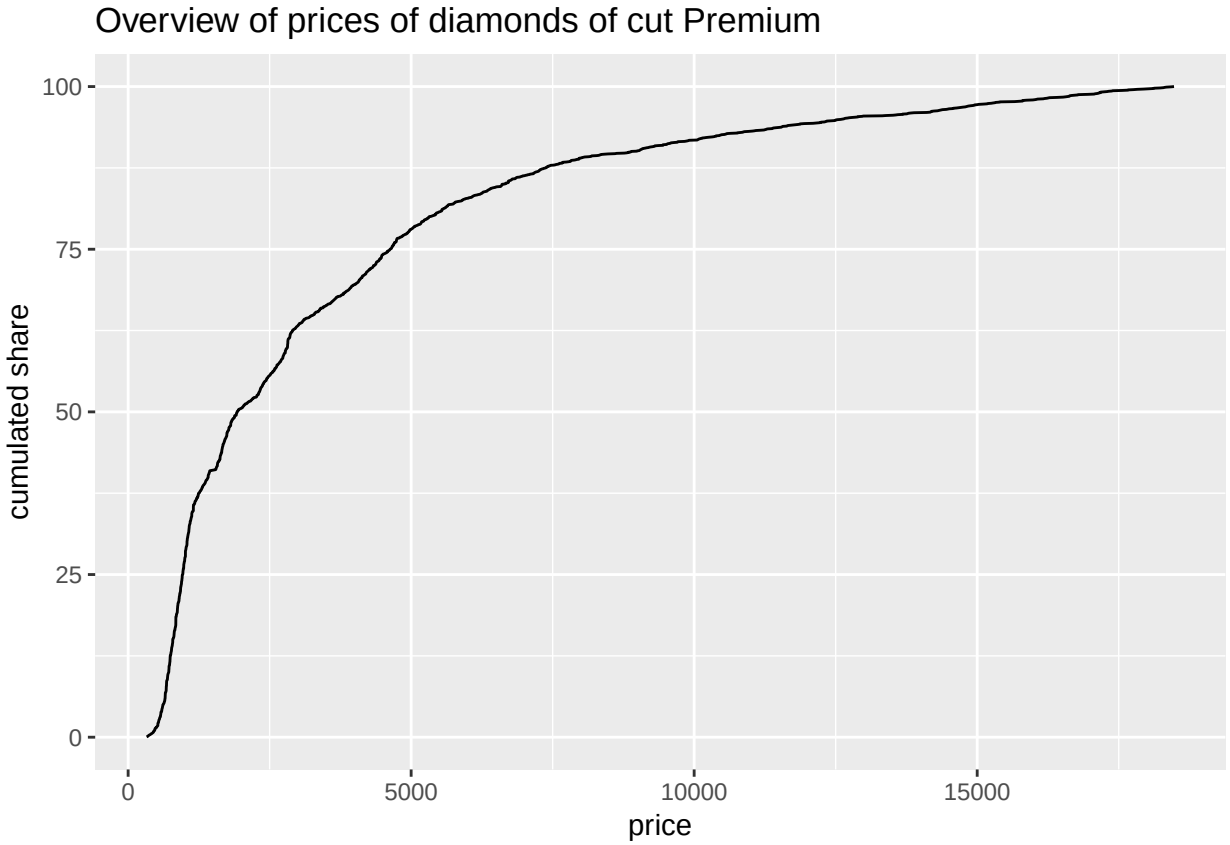
```
  out.loop<-NULL #creating the output

#Creating a chapter for every cut

foreach(y= c("Premium","Fair")) %do% {

  #Parameters used within the loop can be created

  cut.short <- case_when(y=="Ideal"~"I",
                         y=="Premium"~"P",
                         y=="Good"~"G",
                         y=="Very Good"~"V",
                         y=="Fair"~"F"
                         )
  headline<-  paste0("# Section about the cut ", y)

  #Creating the chapter for every cut

  res.loop<-c(headline
      ,   knit_child('loop_child_1.Rmd' , envir=knit_global())
      ,   knit_child('loop_child_2.RMD' , envir=knit_global())
                    )

  #Puting the chapters together

  out.loop <- c(out.loop,res.loop)

  }
```

```
#Pasting the chapters into the text with the following inline code

# `r  paste(out.loop, collapse='\n')`
```

7 Section about the cut Premium

7.1 First child-script

```
data %>%
  filter(substr(cut,1,1)==cut.short) %>%
  group_by(price) %>%
  summarise(nu=n()) %>%
  arrange(price) %>%
  mutate(per=cumsum(nu)/sum(nu)*100) %>%
  ggplot()+geom_line(aes(x=price,y=per)) +
  labs(title = paste0("Overview of prices of diamonds of cut ",y),
       y="cumulated share")
```



7.2 Second child-script

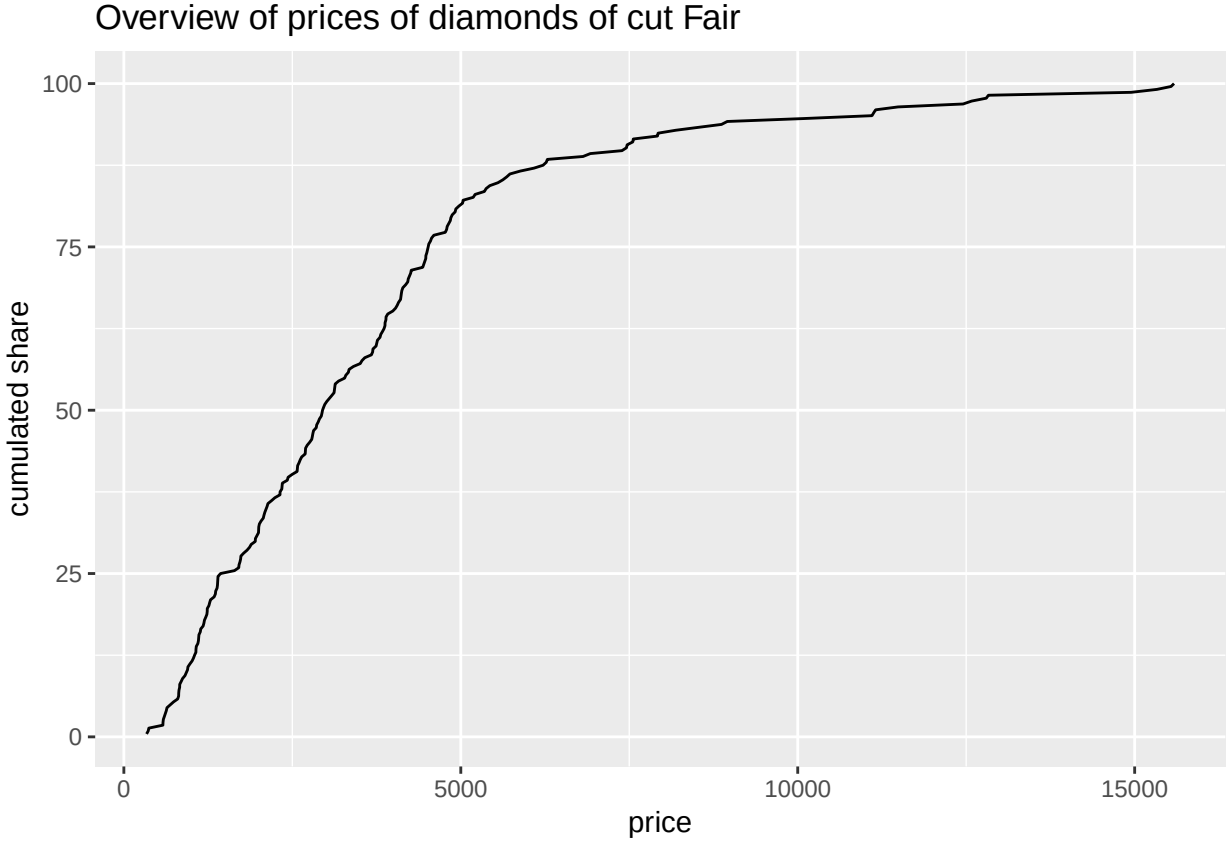
```
t<-
data %>%
  filter(substr(cut,1,1)==cut.short&price>=16000) %>%
  summarise(t=n()) %>% pull(t)
```

48 Diamonds of the cut Premium have a price higher than 16.000.

8 Section about the cut Fair

8.1 First child-script

```
data %>%
  filter(substr(cut,1,1)==cut.short) %>%
  group_by(price) %>%
  summarise(nu=n()) %>%
  arrange(price) %>%
  mutate(per=cumsum(nu)/sum(nu)*100) %>%
  ggplot()+geom_line(aes(x=price,y=per)) +
  labs(title = paste0("Overview of prices of diamonds of cut ",y),
       y="cumulated share")
```



8.2 Second child-script

```
t<-
data %>%
```

```
  filter(substr(cut,1,1)==cut.short&price>=16000) %>%
  summarise(t=n()) %>% pull(t)
```

0 Diamonds of the cut Fair have a price higher than 16.000.