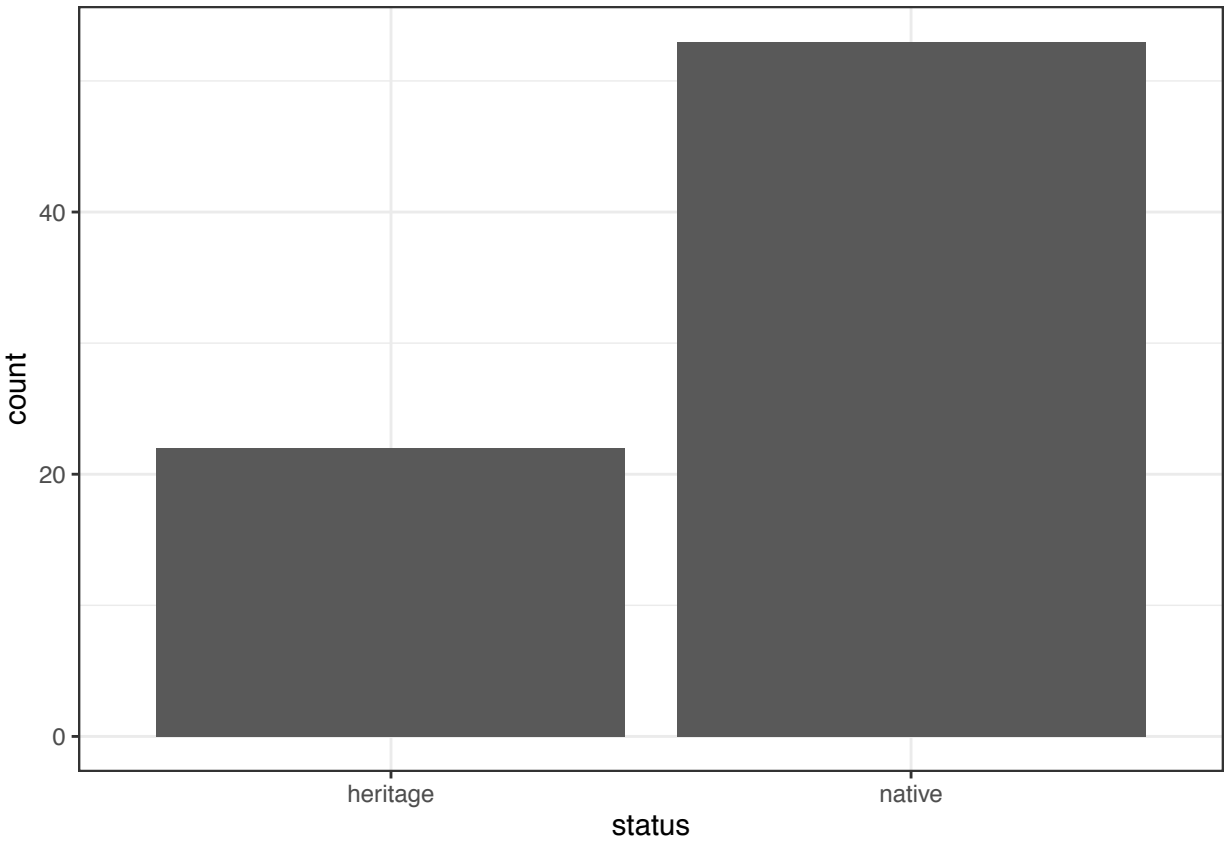
BCS_RSA_Noun_Norming_Native

Participant Demographics

Language status

Foreign: first language and language spoken at school are not BCS
Heritage: first language is BCS and language spoken at school is not BCS
SIMK: Country of origin is Slovenia (SI) or Macedonia (MK)
native: first language and language spoken at school are BCS

For this iteration, I got rid of all SIMK and foreign speakers (their answers made the data a lot more messy)



Dialect Information

Dialect Measure 1

Dialects are often split up based on the phonological change of the Common Slavic *jat* vowel (*ě), which changed to /i/ (Ikavian/Ikavica), /e/ (Ekavian/Ekavica), or /ije/ /je/ (Ijekavian/Ijekavica). Standard

Croatian and Bosnian is based on Ijekavian, whereas standard Serbian is based on Ekavian. Serbians in Croatia, Bosnian Serbs, and Montenegrians mainly use Ijekavian. A geographic distribution is shown below.

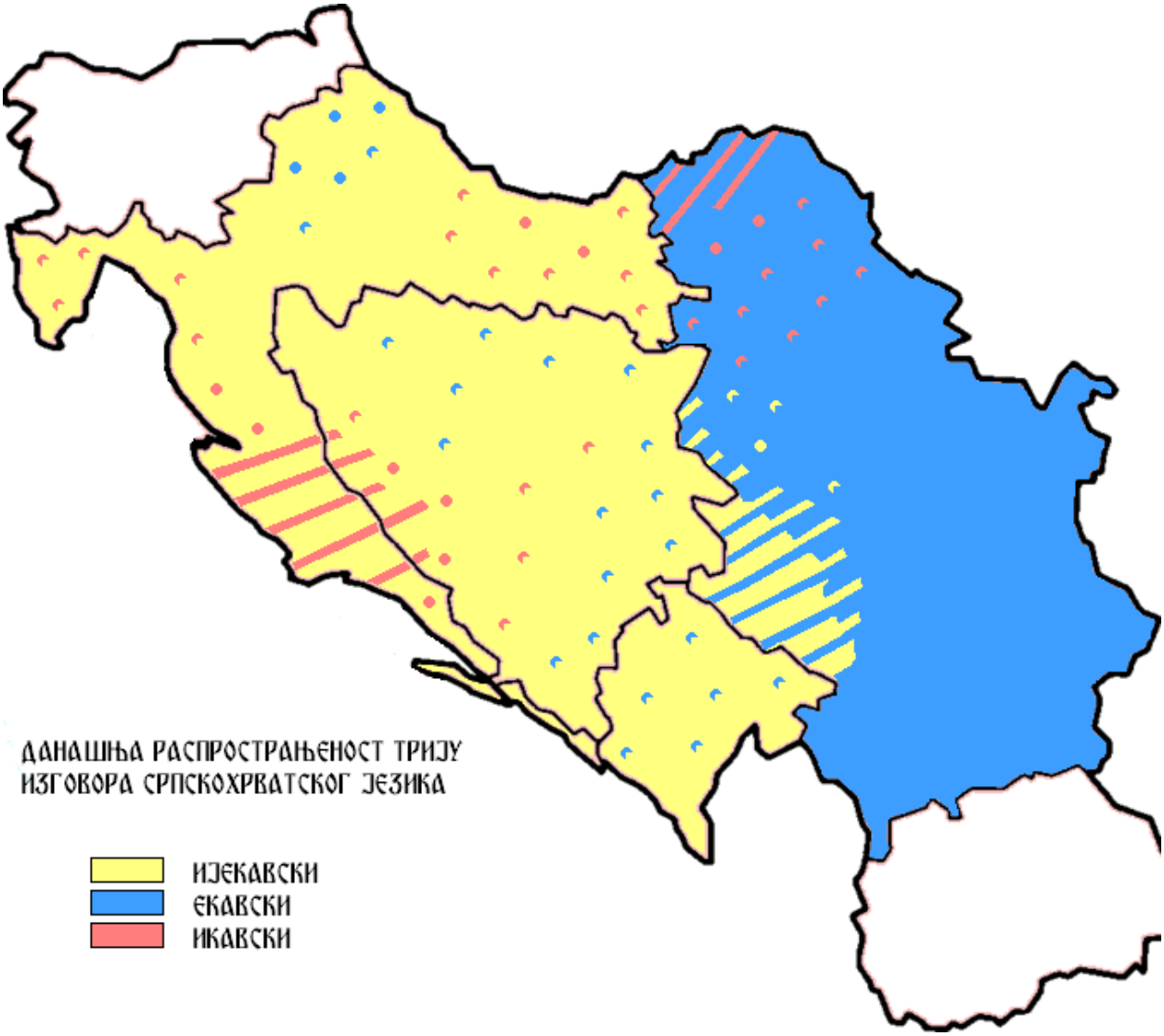
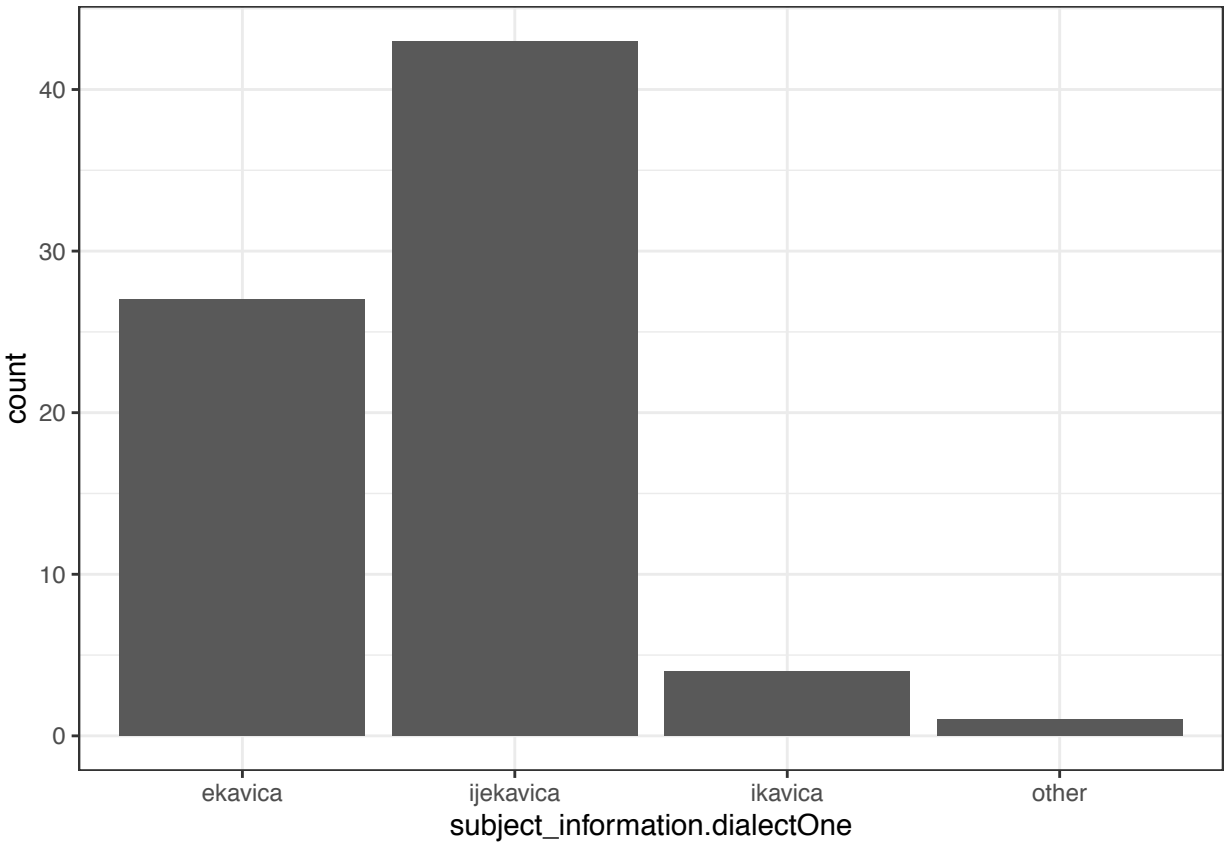


Figure 1: alt text here

yellow: ijekaciva
blue: ekavica
red: ikavica

Our participants had the following distribution:



Dialect Measure 2

Another common way of classifying the dialects is based on their most common wuestion word for "what":

Shtokavian (Štokavski) = što/šta
Chakavian (Čakavski) = Ča/ca
Kajkavian (Kajkavski) = Kaj/key

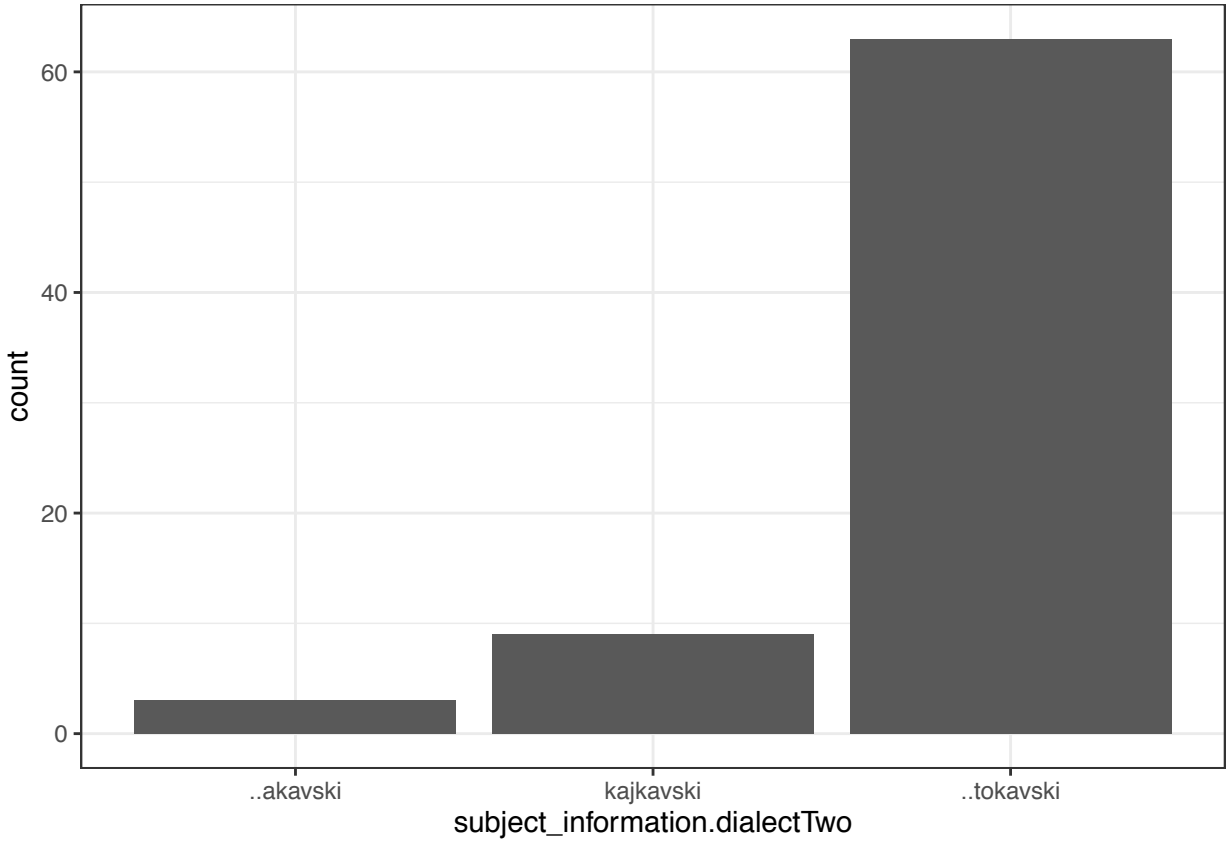
Shtokavian is the standard for all of BCS and is spoken in Bosnia and Herzegovina, Croatia, Montenegro, and Serbia. Chakavian is spoken on the Croatian coast and Kajkavian is spoken in northern Croatia. Slovenians use the word "kaj" for the "what" question word, therefore I assume Slovenians would be classified as Kaikavian speakers based on my dialect survey (even if they don't produce the word "kaj" when speaking BCS).

The map below gives a geographic distribution of the three dialects in Croatia. Bosnia and Herzegovina, Mongenegro, and Serbia would all be colored green based on this map.

Chakavian (blue) = Čakavski
Kaikavian (purple) = Kajkavski
Shtokavian (green) = Štokavski

Our participants had the following dialect distribution:

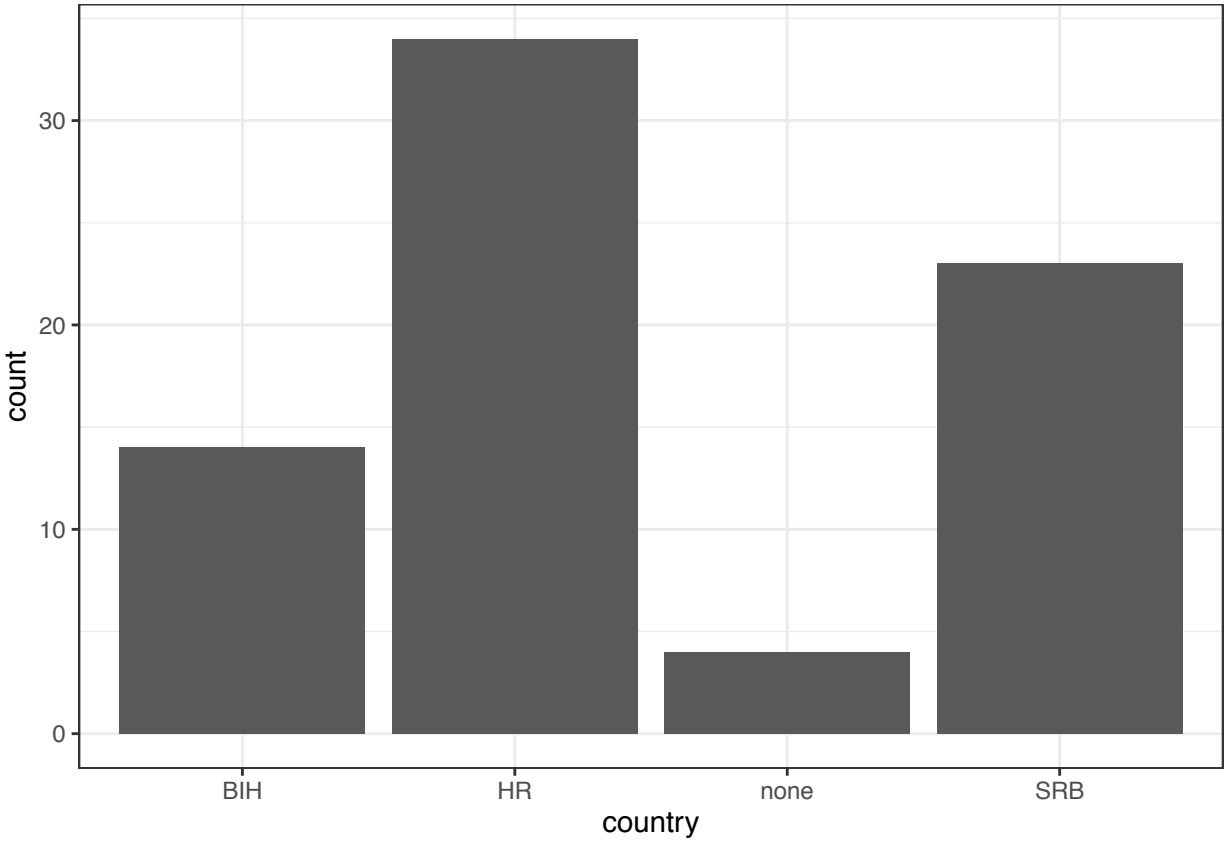
Figure 2: alt text here



Country of participants

Participants answered the question: In which Ex-Jugoslav country did I live in/spent a lot of time in. This question aims to get a geographic distribution of participants. While we can classify speakers across ekavian/ikavian/ijekavian dialects, since this distiction represents a phonological one, it is unclear whether it will likewise reflect a difference in use of lexical items. Keep in mind that the answer to this question is distinct from what nationality participants identify as (for example there are Serbians, Croatians, and Bosnians who all live in Bosnia and Hercegovina). Based on my (limited) observations, I believe that country may be a better predictor of lexical items than national identity could be.

BIH: Bosnia and Herzegovina
HR: Croatia
SRB: Serbia
SI: Slovenia
MK: Macedonia
MNE: Montenegro



Analysis of which words should be used

Stefan and Brandon had a long discussion on how we should choose these items in a principled manner. I will try to recap some of the discussion here and then go through and pick items.

The main fear is the following: we are investigating whether participants use gender pragmatically to modulate how often and when they use nouns versus adjectives redundanly. The question Brandon brought up was, if an object has two lexical items associated with it of two different genders, the participant could potentially make the choice between the two lexical items based on its gender rather than semantic considerations.

Let's assume the following two scenarios: scenario 1: [present_object, shoe_object] scenario 2: [present_object, box_object]

and the following results of teh norming study: present_object: "present_masc" = 40% "box_fem" = 60%

box_object: "box_fem" = 99% "container_masc" = 1%

shoe_object: "shoe_fem" = 98% "sneaker_fem" = 2%

We assume that the results of the norming study give a fair prediction of how likely participants are to use a label given an object, disregarding context.

Clearly context does play a role. For example in scenario 2, participants will opt to call the present_object "present_masc" due to the context.

However, most cases will end up being like scenario 1, whereby a participant can use both "present_masc" or "box_fem". Both will semantically disambiguate the present_object from the shoe_object well enough.

We have to consider the possibility that in scenario 2, gender can play a role in picking the lexical item. So in this scenario, the prior for picking the lexical item "shoe_fem" is really strong. It is a possibility that participants can then reason about the gender of the lexical item they have to pick for the present_object. That is, in scenario 2 they might opt to pick "present_masc" more often than "box_fem" because it will disambiguate the object not just by lexical item but also gender.

I believe this kind of reasoning is unlikely, but as Brandon brought up, since we are looking at whether participants use gender pragmatically to modulate redundant uses of nouns and adjectives—there is no reason why this possibility doesn't also exist.

Therefore I propose the following classification of items: 1. Fall above the 70% threshold –> we can include these items 2. Almost fall above the threshold, and would if I clean up spelling errors and stuff like that –> I will clean it up and we will see if they do fall above the threshold –> we can include these items 3. Fall below threshold, but lexical items used are of the same gender and when added up they reach teh 70% threshold –> In that case we don't have to worry about the issue above, because gender can never be the deciding factor between those two lexical items. –> we can include these items 4. Fall below threshold, and the lexical items used are of different genders. However, the lexical item used is based on dialect. In that case, we have reasonable assumption that we won't run into the issue discussed above, because a single speaker will only use the lexical item (and thus a single gender) given the speach community they are in. (They might have both lexical items if they are bidialectal, but in that case will only presumably use the lexical item in the dialect they are currently communicating in) –> we can include these items 5. Else case –> we can't include these items.

The following table is the % of people that answered the most common answer for each of the following words:

	item	colorScheme	gender	label	percent
1	airplane	2	M	avion	0.9733333
2	armchair	1	F	fotelja	0.8000000
3	balloon	1	M	balon	1.0000000
4	bandaid	2	M	flaster	0.7333333
5	basket	2	F	korpa	0.5200000
6	bed	1	M	krevet	1.0000000
7	belt	1	M	kais	0.6266667
8	bench	2	F	klupa	0.9866667
9	bike	1	M	bicikl	0.5466667
10	billiardball	1	F	kugla	0.3600000
11	binder	1	F	fascikla	0.3466667
12	binoculars	2	M	dvogled	0.5866667
13	bonbon	2	F	bombon	0.4933333
14	book	1	F	knjiga	0.9200000
15	boot	2	F	cizma	0.6666667

```
## 16             bowl          2      F       zdjela 0.5866667
## 17              box          2      F       kutija 0.9200000
## 18        bracelet           1      F    narukvica 0.7066667
## 19          bucket           1      F        kanta 0.7600000
## 20       butterfly           1      M       leptir 1.0000000
## 21            cake           2      F        torta 0.9466667
## 22      calculator           1      M     digitron 0.4533333
## 23        calendar           2      M     kalendar 0.9866667
## 24          camera           2      F   fotoaparat 0.5333333
## 25          candle           1      F        sveca 0.9200000
## 26             cap           1      M         kapa 0.5200000
## 27           chair           1      F      stolica 0.9066667
## 28           clock           1      M          sat 0.3866667
## 29      coathanger           1      M     vesalica 0.6133333
## 30            comb           1      M       cesalj 0.9466667
## 31           crown           1      F        kruna 0.9866667
## 32         cushion           1      M       jastuk 0.9600000
## 33             die           2      F        kocka 1.0000000
## 34            door           1      F        vrata 1.0000000
## 35           dress           1      F      haljina 0.9733333
## 36         dresser           2      F       komoda 0.4000000
## 37            drum           2      M       bubanj 0.8933333
## 38            duck           2      F        patka 0.5733333
## 39         dustpan           1      M     lopatica 0.3333333
## 40           fence           1      F       ograda 0.8533333
## 41            fish           2      F         riba 0.9600000
## 42          flower           1      M       cvijet 0.5733333
## 43            fork           2      F     viljuska 0.5333333
## 44           frame           1      M         okvir 0.5466667
## 45       fryingpan           1      M         tava 0.6266667
## 46           glove           2      F     rukavica 0.8933333
## 47          guitar           1      F       gitara 0.8533333
## 48          hammer           2      M        cekic 0.9466667
## 49      helicopter           2      M   helikopter 0.8800000
## 50            iron           1      F        pegla 0.8800000
## 51             key           1      M        kljuc 0.9200000
## 52           knife           2      M          noz 0.6933333
## 53           ladle           2      F      kutlaca 0.4666667
## 54            lamp           1      F        lampa 0.7466667
## 55       lipstick            2      M       karmin 0.4133333
## 56            lock           2      M      katanac 0.4533333
## 57         luggage           2      M        kofer 0.7733333
## 58  magnifyingglass           1      F     povecalo 0.4666667
## 59            mask           1      F        maska 0.8533333
## 60      microscope           2      M    mikroskop 0.8800000
## 61           mouse           1      M          mis 0.6133333
## 62             mug           1      F        solja 0.4533333
## 63          napkin           1      F      salveta 0.4666667
## 64        necklace           2      F      ogrlica 0.7866667
## 65        notebook           1      F       sveska 0.3200000
## 66        ornament           1      F kuglicazabor 0.1466667
## 67          pencil           1      F       olovka 0.8933333
## 68           phone           1      M      telefon 0.9866667
## 69           piano           2      M       klavir 0.9200000
```

```
## 70            plate           2      M       tanjir 0.6133333
## 71          present           1      M       poklon 0.6800000
## 72            purse           2      F        torba 0.4666667
## 73            radio           2      M        radio 0.6666667
## 74            razor           2      M       brijac 0.4933333
## 75           remote           2      M    daljinski 0.5200000
## 76           ribbon           2      F        masna 0.8266667
## 77             ring           1      M       prsten 0.8666667
## 78            robot           2      M        robot 0.8400000
## 79             rope           2      M      konopac 0.2266667
## 80              rug           1      M        tepih 0.6000000
## 81            ruler           1      M      ravnalo 0.4133333
## 82            scarf           1      M          sal 0.6533333
## 83      screwdriver           2      M    srafciger 0.2800000
## 84            shell           2      F      skoljka 0.8533333
## 85           shield           2      M         stit 0.8400000
## 86            shirt           1      F       majica 0.7066667
## 87             shoe           1      F       cipela 0.6933333
## 88           shovel           2      F       lopata 0.7333333
## 89          slipper           2      F       papuca 0.7466667
## 90             sock           2      F       carapa 0.8533333
## 91           sponge           2      M       spuzva 0.3600000
## 92            spoon           2      F       kasika 0.5333333
## 93          stapler           1      F    heftalica 0.2800000
## 94           switch           1      M     prekidac 0.6533333
## 95            sword           1      M          mac 0.4266667
## 96            table           1      M        klupa 0.5066667
## 97             tent           2      M        sator 0.9066667
## 98              tie           2      F      kravata 0.9066667
## 99            truck           2      M       kamion 0.9200000
## 100        umbrella           2      M     kisobran 0.9733333
## 101            vase           2      F         vaza 0.6800000
## 102          wallet           1      M     novcanik 0.9733333
## 103         whistle           2      F   zvizdaljka 0.4666667
## 104            yarn           1      F         vuna 0.4400000
```

The way we designed the study, we want a total of 48 targets per participant (+ 22 filler trials). That is:
4 scenarios x 2 gender for target x 6 = 48 target trials
+ 22 filler trials
= 70 trials per pariticipant.

There will be two stimuli sets based on color scheme:
colorscheme 1 (cs1 = {blue, yellow, red, white}; cs2 = {green, purple, orange, black}
For each trial, the stimuli images will be drawn randomly from within a single color scheme. Therefore we need a total of 96 items (each in 4 colors). Their distribution is as follows:

96 items:
- 48 in color scheme 1
- 24 masculine
- 24 feminine
- 48 in color scheme 2
- 24 masculine
- 24 feminine

In order to see if we have enough stimuli based on our nameability norming, I group all the stimuli based on their color scheme and gender and then sort the nouns into the following groups:

1. > 70% of responses are the same
2. <70% of responses are the same, but if we account for spelling errors it seems that we will get >70% (I have yet to go through and do this manually)
3. there are 2-3 majority words, the object is clearly identifiable, and all the words are the same gender
4. there are 2-3 majority words, the object is clearly identifiable, but the words are of different genders
5. there is no majority word and the object is clearly identifiable
6. there is no majority word and the object is not clearly identifiable

Overall the distribution of words with >70% responses agree on a single response are:

```
##   n()
## 1  52
```

The number of responses that do not meet the 70% threshold are:

```
##   n()
## 1  52
```

We now split up by color scheme:

Color Scheme 1

The number of responses >= 70% threshold in color scheme 1 split up by gender:

```
## # A tibble: 2 x 2
##   gender `n()`
##   <fct>  <int>
## 1 F         16
## 2 M          9
```

Feminine

Here are all the items in color scheme one that are feminine

```
##               item colorScheme gender        label   percent
## 1         ornament           1      F kuglicazabor 0.1466667
## 2          stapler           1      F    heftalica 0.2800000
## 3         notebook           1      F       sveska 0.3200000
## 4           binder           1      F     fascikla 0.3466667
## 5     billiardball           1      F        kugla 0.3600000
## 6             yarn           1      F         vuna 0.4400000
## 7              mug           1      F        solja 0.4533333
## 8  magnifyingglass           1      F     povecalo 0.4666667
## 9           napkin           1      F      salveta 0.4666667
## 10            shoe           1      F       cipela 0.6933333
## 11        bracelet           1      F    narukvica 0.7066667
## 12           shirt           1      F       majica 0.7066667
## 13            lamp           1      F        lampa 0.7466667
## 14          bucket           1      F        kanta 0.7600000
## 15        armchair           1      F      fotelja 0.8000000
## 16           fence           1      F       ograda 0.8533333
## 17          guitar           1      F       gitara 0.8533333
## 18            mask           1      F        maska 0.8533333
## 19            iron           1      F        pegla 0.8800000
## 20          pencil           1      F       olovka 0.8933333
## 21           chair           1      F      stolica 0.9066667
```

```
## 22            book           1      F       knjiga 0.9200000
## 23          candle           1      F        sveca 0.9200000
## 24           dress           1      F      haljina 0.9733333
## 25           crown           1      F        kruna 0.9866667
## 26            door           1      F        vrata 1.0000000
```

As stated above, there are 16 feminine nouns that fall into category (1)

Fall above the 70% threshold
- door
- crown
- dress - candle
- book
- chair
- pencil
- iron
- mask
- guitar
- fence
- armchair
- bucket
- lamp
- shirt
- bracelet

Here are the words that fell below the threshold:

```
##               item colorScheme gender        label   percent
## 1         ornament           1      F kuglicazabor 0.1466667
## 2          stapler           1      F    heftalica 0.2800000
## 3         notebook           1      F       sveska 0.3200000
## 4           binder           1      F     fascikla 0.3466667
## 5     billiardball           1      F        kugla 0.3600000
## 6             yarn           1      F         vuna 0.4400000
## 7              mug           1      F        solja 0.4533333
## 8  magnifyingglass           1      F     povecalo 0.4666667
## 9           napkin           1      F      salveta 0.4666667
## 10            shoe           1      F       cipela 0.6933333
```

Words that fall above the threshold if we combine the top words with the same gender:
- shoe: {stikla, cipela}
- mug: {casa, salica, solja} - stapler: {heftalica, klamerica, spajalica}

Words that we should include contingent on dialect:
- napkin: {marama, krpa, salveta} for SRB and BIH - magnifyingglass: {lupa} for SRB

words to consider due to compounds: - billiardball

Words we should not include:
- yarn
- binder
- notebook

Overall we should include: all words above threshold, all words that are of the same gender when combined fall above threshold
This gives us 19 target words

If we also want to include words contingent on dialect: this vies us: 21 target words

Shoe

```
## # A tibble: 14 x 2
## # Groups:   response [14]
##    response             n
##    <chr>            <int>
##  1 baletanka            1
##  2 cipela              52
##  3 cipelanastiklu       1
##  4 cipelasalonka        1
##  5 cipelaspetom         1
##  6 cipelespetom         1
##  7 obuca                1
##  8 papuca               1
##  9 postola              1
## 10 sivecipelenapetu     1
## 11 stijlke              1
## 12 stikla              11
## 13 stikla/cipela        1
## 14 zipela               1
```

Lets combine the top answers “stikla” (high hell) and “cipela” (shoe)

```
##           n
## 1 0.8666667
```

Combining “stikla” and “cipela” gives us the threshold –> type 3, should include

Napkin

```
## # A tibble: 18 x 2
## # Groups:   response [18]
##    response                      n
##    <chr>                     <int>
##  1 bijelirupcic                  1
##  2 guma                          1
##  3 kirpa                         1
##  4 krpa                          5
##  5 marama                        1
##  6 maramcia,krpa                 1
##  7 maramica                     12
##  8 neznam                        1
##  9 neznam,mozdakapailiubrus?     1
## 10 obrusladja                    1
## 11 papir                         2
## 12 rupcic                        1
## 13 salveta                      35
## 14 salveta/krpa                  1
## 15 sarvet                        1
## 16 savleta                       1
## 17 selveta                       1
## 18 ubrus                         8
```

Lets combine the top answers “salveta” (napkin), “marama” (cloth for clothing), “krpa” (rag)

```
##           n
## 1 0.7333333
```

However we should consider that `0.133333333333333` of participants answered some variant of "ubrus". Is that a dialect thing?

```
## # A tibble: 16 x 3
## # Groups:   subject_information.dialectOne, response [16]
##    subject_information.dialectOne response         n
##    <fct>                          <chr>        <int>
##  1 ekavica                        kirpa            1
##  2 ekavica                        majority        23
##  3 ekavica                        neznam           1
##  4 ekavica                        salveta          1
##  5 ekavica                        ubrus            1
##  6 ijekavica                      bijelirupcic     1
##  7 ijekavica                      majority        29
##  8 ijekavica                      papir            2
##  9 ijekavica                      rupcic           1
## 10 ijekavica                      salveta          1
## 11 ijekavica                      savleta          1
## 12 ijekavica                      ubrus            7
## 13 ijekavica                      ubrus?           1
## 14 ikavica                        majority         3
## 15 ikavica                        ubrus            1
## 16 other                          guma             1
```

```
## # A tibble: 16 x 3
## # Groups:   subject_information.dialectTwo, response [16]
##    subject_information.dialectTwo response         n
##    <fct>                          <chr>        <int>
##  1 čakavski                       majority         2
##  2 čakavski                       ubrus            1
##  3 kajkavski                      guma             1
##  4 kajkavski                      majority         6
##  5 kajkavski                      salveta          1
##  6 kajkavski                      ubrus            1
##  7 štokavski                      bijelirupcic     1
##  8 štokavski                      kirpa            1
##  9 štokavski                      majority        47
## 10 štokavski                      neznam           1
## 11 štokavski                      papir            2
## 12 štokavski                      rupcic           1
## 13 štokavski                      salveta          1
## 14 štokavski                      savleta          1
## 15 štokavski                      ubrus            7
## 16 štokavski                      ubrus?           1
```

```
## # A tibble: 16 x 3
## # Groups:   country, response [16]
##    country response         n
##    <fct>   <chr>        <int>
##  1 BIH     majority        10
##  2 BIH     papir            1
##  3 BIH     salveta          1
##  4 BIH     savleta          1
##  5 BIH     ubrus?           1
##  6 HR      bijelirupcic     1
```

```
##  7 HR      guma             1
##  8 HR      majority        21
##  9 HR      papir            1
## 10 HR      rupcic           1
## 11 HR      ubrus            9
## 12 none    majority         3
## 13 none    salveta          1
## 14 SRB     kirpa            1
## 15 SRB     majority        21
## 16 SRB     neznam           1
```

Of the dialect measures, what seems telling is 1 instance was from BIH and all the other 9 were from Croatia. Therefore this might be a dialectal measure which features for Croat speakers.

Croatians:

```
##           n
## 1 0.6176471
```

BIH:

```
##           n
## 1 0.7142857
```

SRB:

```
##           n
## 1 0.9130435
```

The picture is complex. The combination of "salveta", "marama", and "krpa" reaches threshold for Serbia and BIH, but not Croatia. Most Ubrus instances are found for speakers in HR and who speak Ijekavica (the dialect in BIH and HR). Therefore if we include this word, we should look at whether HR speakers in the RSA study use ubrus, and if so, exclude all of them from the analysis for this word.

Magnifyingglass

```
## # A tibble: 10 x 2
## # Groups:   response [10]
##    response              n
##    <chr>             <int>
##  1 lupa                 32
##  2 lupa/povecalo         1
##  3 lupo                  1
##  4 magnifyingglas        1
##  5 mikroskop             1
##  6 neznam                1
##  7 ogledalo              1
##  8 povecalo             35
##  9 staklozapovecanje     1
## 10 uvecalo               1
```

```
## # A tibble: 10 x 2
## # Groups:   response [10]
##    response              n
##    <chr>             <int>
##  1 lupa                 32
##  2 lupa/povecalo         1
##  3 lupo                  1
```

```
##  4 magnifyingglas        1
##  5 mikroskop             1
##  6 neznam                1
##  7 ogledalo              1
##  8 povecalo             35
##  9 staklozapovecanje     1
## 10 uvecalo               1
```

```
## # A tibble: 14 x 3
## # Groups:   subject_information.dialectOne, response [14]
##    subject_information.dialectOne response             n
##    <fct>                          <chr>             <int>
##  1 ekavica                        lupa                 24
##  2 ekavica                        povecalo              2
##  3 ekavica                        uvecalo               1
##  4 ijekavica                      lupa                  8
##  5 ijekavica                      lupa/povecalo         1
##  6 ijekavica                      lupo                  1
##  7 ijekavica                      magnifyingglas        1
##  8 ijekavica                      mikroskop             1
##  9 ijekavica                      neznam                1
## 10 ijekavica                      ogledalo              1
## 11 ijekavica                      povecalo             28
## 12 ijekavica                      staklozapovecanje     1
## 13 ikavica                        povecalo              4
## 14 other                          povecalo              1
```

```
## # A tibble: 15 x 3
## # Groups:   country, response [15]
##    country response             n
##    <fct>   <chr>             <int>
##  1 BIH     lupa                  5
##  2 BIH     lupa/povecalo         1
##  3 BIH     lupo                  1
##  4 BIH     magnifyingglas        1
##  5 BIH     neznam                1
##  6 BIH     povecalo              5
##  7 HR      lupa                  4
##  8 HR      mikroskop             1
##  9 HR      ogledalo              1
## 10 HR      povecalo             27
## 11 HR      staklozapovecanje     1
## 12 none    lupa                  1
## 13 none    povecalo              3
## 14 SRB     lupa                 22
## 15 SRB     uvecalo               1
```

HR:

```
##           n
## 1 0.1176471
```

BIH:

```
##           n
## 1 0.3571429
```

SRB:

```
##           n
## 1 0.9565217
```

Magnifying glass as "lupa" only reaches the threshold for serbian speakers. If we include it, we should only include the data from Serbian sepaekrs and not people from HR or BIH.

Mug

```
## # A tibble: 7 x 2
## # Groups:   response [7]
##   response        n
##   <chr>       <int>
## 1 casa            8
## 2 casazakavu      1
## 3 cikara          1
## 4 plavasalica     1
## 5 salica         29
## 6 soju            1
## 7 solja          34
```

The confusion here is between: "casa" (cup), "salica" (mug), and "solja" (mug)

```
##           n
## 1 0.9733333
```

All three words are of the same gender and reach threshold. We should include it.

Yarn

```
## # A tibble: 21 x 2
## # Groups:   response [21]
##    response           n
##    <chr>          <int>
##  1 crvenavuna         1
##  2 garnislo           1
##  3 klupko            11
##  4 klupkovune         5
##  5 kolut              1
##  6 konac              4
##  7 konopac            1
##  8 nit.               1
##  9 niz                1
## 10 odcesljanevune     1
## # ... with 11 more rows
```

Use of "vuna" (yarn) when accounting for different spellings:

```
##           n
## 1 0.5466667
```

This one is very messy in terms of answers, I think we should not include this one.

billiardball

```
## # A tibble: 25 x 2
## # Groups:   response [25]
##    response            n
##    <chr>           <int>
##  1 bilijardkugla       1
##  2 bilijarksakugla     1
##  3 bilijarkugla        1
##  4 bilijarskakugla     5
##  5 bilijarskalopta     2
##  6 biljardkugla        5
##  7 biljarkugla         2
##  8 biljarlopta         1
##  9 biljarskalopta      1
## 10 billijarda          1
## # ... with 15 more rows
```

If we count all compounds with "lopta" (ball) and kugla (ball):

```
##      n
## 1 0.88
```

We get a clear majority. Include this one if we want to include compounds.

If we count all compounds with "lopta" (ball) and kugla (ball):

```
##      n
## 1 0.88
```

We get a clear majority. Include this one if we want to include compounds.

binder

```
## # A tibble: 24 x 2
## # Groups:   response [24]
##    response           n
##    <chr>          <int>
##  1 dnevnik            1
##  2 dosije             1
##  3 fascikl           11
##  4 fascikla          26
##  5 fasckil            1
##  6 filofaks           1
##  7 folder             9
##  8 futrola            1
##  9 knigu              1
## 10 knjigovezacn/a     1
## # ... with 14 more rows
```

This one is a mess, we shouldn't include it.

notebook

```
## # A tibble: 26 x 2
## # Groups:   response [26]
##    response            n
##    <chr>           <int>
##  1 beleznica          21
```

```
##  2 beleznica/teka      1
##  3 biljaznica          1
##  4 biljeska            1
##  5 biljeske            1
##  6 binder              1
##  7 ctirtanjeknjiga     1
##  8 dnjevnik            1
##  9 fascikl             1
## 10 fascikla            1
## # ... with 16 more rows
```

This one is a mess, we shouldn't include it.

stapler

```
## # A tibble: 16 x 2
## # Groups:   response [16]
##    response                                        n
##    <chr>                                       <int>
##  1 drukerica                                       1
##  2 heftalica                                      21
##  3 heftarica                                       5
##  4 hektalica                                       1
##  5 klamaric                                        1
##  6 klamarica                                       5
##  7 klamerica                                      15
##  8 klemerica                                       2
##  9 nemampojma-nikadganisamkoristilaujugoslaviji    1
## 10 nemoguaesjetiti                                 1
## 11 nesjecamse                                      1
## 12 neznam                                          5
## 13 spajalica                                      13
## 14 stampa                                          1
## 15 stapler                                         1
## 16 stempler                                        1
```

Majority words are: spajalica, heftalica, and klamerica. All are of the same gender. If we combine these responses we get:

```
##           n
## 1 0.8266667
```

If we combine all 3, they reach threshold. I assume the three words come from different dialects, but since they are all the same gender it doesn't really matter. We should include it.

Ornament

```
## # A tibble: 28 x 2
## # Groups:   response [28]
##    response                n
##    <chr>                <int>
##  1 bijelakuglicazaboric     1
##  2 bolba                    1
##  3 bozicnakugla             3
##  4 bozicnakuglica           1
##  5 bozicnalopta             1
```

```
##  6 bozicniukras             1
##  7 cuglice                  1
##  8 duvetlopta               1
##  9 kiglica                  1
## 10 kugla                    8
## # ... with 18 more rows
```

People understood what the object was but the responses were a mess and included a lot of compounds. I don't think we should include this one.

Masculine CS 1

Here are all the items in color scheme one that are masculine

```
##          item colorScheme gender    label   percent
## 1     dustpan           1      M lopatica 0.3333333
## 2       clock           1      M      sat 0.3866667
## 3       ruler           1      M  ravnalo 0.4133333
## 4       sword           1      M      mac 0.4266667
## 5  calculator           1      M digitron 0.4533333
## 6       table           1      M    klupa 0.5066667
## 7         cap           1      M     kapa 0.5200000
## 8        bike           1      M   bicikl 0.5466667
## 9       frame           1      M    okvir 0.5466667
## 10     flower           1      M   cvijet 0.5733333
## 11        rug           1      M    tepih 0.6000000
## 12 coathanger           1      M vesalica 0.6133333
## 13      mouse           1      M      mis 0.6133333
## 14       belt           1      M     kais 0.6266667
## 15  fryingpan           1      M     tava 0.6266667
## 16      scarf           1      M      sal 0.6533333
## 17     switch           1      M prekidac 0.6533333
## 18    present           1      M   poklon 0.6800000
## 19       ring           1      M   prsten 0.8666667
## 20        key           1      M    kljuc 0.9200000
## 21       comb           1      M   cesalj 0.9466667
## 22    cushion           1      M   jastuk 0.9600000
## 23     wallet           1      M novcanik 0.9733333
## 24      phone           1      M  telefon 0.9866667
## 25    balloon           1      M    balon 1.0000000
## 26        bed           1      M   krevet 1.0000000
## 27  butterfly           1      M   leptir 1.0000000
```

As stated above, there are 10 masculine nouns that fall into category (1)

Fall above the 70% threshold
- balloon
- bed
- phone
- key
- butterfly
- cushion
- wallet
- comb

- ring
- scarf

Words that fall above threshold after accounting for spelling errors: - scarf - flower

Words that fall above the threshold if we combine the top words with the same gender:
- belt: {kais, remen} - calculator: {digitron, kalkulator}

Words that we should include contingent on dialect:
- fryingpan: {tiganj vs. tava} for SRB vs BIH/HR - ruler: {lenjir} for SRB/BIH, {ravnalo} for HR - clock: {sat} everyone, {budilnik} for SRB, {budilica} for HR, unclear for BIH

words to consider due to compounds: - switch (68% without compounds): {osigurac, prekidac, salter} - mouse (67% without compounds)

Words we should not include:
- present - coathanger (reaches threshold for HR, see notes) - rug - bike - hat - table - sword

Overall we should include: all words above threshold, all words that are of the same gender when combined fall above threshold
This gives us 14 target words

If we also want to include words contingent on dialect: this gives us: 17 target words

If we also want to include words with compounds: this gives us: 19 target words

Considerations for each words that fell below the 70% threshold are given below

Present

```
## # A tibble: 12 x 2
## # Groups:   response [12]
##    response               n
##    <chr>              <int>
##  1 belakutijazapoklon     1
##  2 dar                    2
##  3 dat                    1
##  4 kutija                 5
##  5 kutijazapoklon         3
##  6 kutijha                1
##  7 packet                 1
##  8 paket                  3
##  9 plavakutijazapoklon    1
## 10 poklon                51
## 11 poklonkutija           5
## 12 prezent                1
```

Consider getting rid of the 2 participants that answered “dar”. This is obviously google translate (“dar” is a “gift”, but not a physical one, rather like a talent).

```
## # A tibble: 10 x 2
## # Groups:   response [10]
##    response          n
##    <chr>         <int>
##  1 compound          9
##  2 dar               2
##  3 dat               1
##  4 kutija            5
##  5 kutijha           1
```

```
##  6 packet            1
##  7 paket             3
##  8 plavacompound     1
##  9 poklon           51
## 10 prezent           1
```

We should not include this one. Kutija (box) and Poklon (present) are in competition, and they are of different genders.

switch

```
## # A tibble: 24 x 2
## # Groups:   response [24]
##    response                    n
##    <chr>                   <int>
##  1 crveniprekidaczasvjetlo     1
##  2 kontakt                     1
##  3 neznam                      2
##  4 okidac                      1
##  5 osigurac                    1
##  6 otikac                      1
##  7 porekidac                   1
##  8 prekidac                   49
##  9 prekidac/salter             1
## 10 prekidaczasvetlo            1
## # ... with 14 more rows
```

Cleanup: prekidac, osigurac, Salter compound

```
##      n
## 1 0.68
```

```
##           n
## 1 0.7066667
```

```
##     n
## 1 0.8
```

Prekidac alone accounts for 68% of responses.

If we include "osigurac" accounts for 71% of the responses.

If we also include compounds like "Prekidac za svetlo" (switch for light) we get 80%.

I think we should include this one. Of the responses that are not masculine we have:
siba (1) F sklopka (1) F
svetlo/svijetlo (2) N
uticnica (1) F

However svetlo/svijetlo means "light", which is obviously not what the object is. Likewise "uticnica" is an outlet, which is likewise not what this object is.

Therefore I think we should include this item, and just get rid of trials where the word is used as a compound.

scarf

```
## # A tibble: 22 x 2
## # Groups:   response [22]
##    response       n
```

```
##    <chr>      <int>
##  1 belisal        2
##  2 bijelisal      1
##  3 crvenisal      1
##  4 esarpa         1
##  5 esarpa/sal     1
##  6 marama         4
##  7 plavisal       1
##  8 sal           49
##  9 sal,sal        1
## 10 sal.           1
## # ... with 12 more rows
```

```
##           n
## 1 0.8266667
```

After corrections, this meets the threshold. We should include it.

fryingpan

```
## # A tibble: 6 x 2
## # Groups:   response [6]
##   response         n
##   <chr>        <int>
## 1 pan              1
## 2 serpa            1
## 3 tanjir           1
## 4 tava            49
## 5 tava,tepsija     1
## 6 tiganj          22
```

Distribution of answers for Ekavica:

```
## # A tibble: 4 x 2
## # Groups:   response [4]
##   response         n
##   <chr>        <int>
## 1 pan              1
## 2 tava             3
## 3 tava,tepsija     1
## 4 tiganj          22
```

This reaches the threshold:

```
## [1] 0.8148148
```

Distribution of answers for Ijekavica:

```
## # A tibble: 3 x 2
## # Groups:   response [3]
##   response     n
##   <chr>    <int>
## 1 serpa        1
## 2 tanjir       1
## 3 tava        41
```

This reaches the threshold:

```
## [1] 0.9534884
```

If we group by country we get the same results:

For Serbia:

```
## [1] 0.826087
```

For Croatia:

```
## [1] 0.9411765
```

For BIH:

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

Tava versus Tiganj is entirely dialectal. Despite being of two different genders, we can still include the word because we can reasonably assume that participants will use the one in their dialect.

belt

```
## # A tibble: 5 x 2
## # Groups:   response [5]
##   response      n
##   <chr>     <int>
## 1 kais         47
## 2 pojas         7
## 3 remen        19
## 4 zbuka         1
## 5 zutiremen     1
```

If we combine kais and remen we get threshold:

```
##      n
## 1 0.88
```

Kais and remen are both of the same gender, so we should include this item.

mouse

```
## # A tibble: 21 x 2
## # Groups:   response [21]
##    response                  n
##    <chr>                 <int>
##  1 crvenikompjutorskimis     1
##  2 crvenmis                  1
##  3 igrackimis                1
##  4 kompjuterskimis           6
##  5 kompterskimis             1
##  6 maos                      1
##  7 maus                      1
##  8 mis                      46
##  9 mis,mis                   1
## 10 mis.                      1
## # ... with 11 more rows
```

If we clean up spelling errors we get:

```
##           n
## 1 0.6666667
```

If we include compounds we get:

```
##           n
## 1 0.9066667
```

We get an overwhelming majority.

I think we should include this word and just throw out trials where people use compounds.

coathanger

```
## # A tibble: 7 x 2
## # Groups:   response [7]
##   response             n
##   <chr>            <int>
## 1 bijelavesalica       1
## 2 civiluk              1
## 3 neznam               1
## 4 ofinger             23
## 5 okvir                1
## 6 vesalica            46
## 7 vesalica/ofinger     2
```

Majority responses are vesalica and ofinger.

Let's see if that is based on dialect: Ekavica:

```
## [1] 0.4814815
```

Ijekavica:

```
## [1] 0.6976744
```

SRB:

```
## [1] 0.4782609
```

HR:

```
## [1] 0.8235294
```

BIH:

```
## [1] 0.4285714
```

While this was classified as a masculine noun, maybe we should include it as a feminine noun and only look at croatian speakers, as that is the only time this reached threshold. However, I believe the majority of our participants will be serbian speakers (based on community outreach), so we might also consider not including this item.

rug

```
## # A tibble: 14 x 2
## # Groups:   response [14]
##    response          n
##    <chr>         <int>
##  1 carpet            1
##  2 cilim             7
##  3 deka/pokrivac     1
##  4 krpara            2
##  5 neznam            1
##  6 otirac            2
##  7 platno            1
```

```
##  8 podmetac          1
##  9 prostirka         4
## 10 rucnik            1
## 11 stazica           2
## 12 tapet             4
## 13 tapison           1
## 14 tepih            47
```

Let's combine the items: cilin, tapet, tepih

```
##      n
## 1 0.68
```

This doesn't reach threshold. Given the semantic differences with the other words, I don't think we should include it.

flower

```
##      n
## 1 0.96
```

After cleaning up spellings, this reaches threshold. We should include it.

frame

```
## # A tibble: 13 x 2
## # Groups:   response [13]
##    response         n
##    <chr>        <int>
##  1 bijeliprozor     1
##  2 ogledalo         2
##  3 okvir           42
##  4 okvirzaslike     2
##  5 okvirzasliku     1
##  6 prozor           1
##  7 ram             15
##  8 ramzaslike       4
##  9 ramzasliku       3
## 10 slika            1
## 11 sliku            1
## 12 steker           1
## 13 tanjir           1
```

If we combine ram and okvir:

```
##      n
## 1 0.76
```

It reaches threshold. Since ram and okvir are the same gender, we should include this word.

bike

```
## # A tibble: 7 x 2
## # Groups:   response [7]
##   response        n
##   <chr>       <int>
## 1 biciclo         1
```

```
## 2 bicika          1
## 3 bicikl         41
## 4 bicikla        13
## 5 bicikli         3
## 6 biciklo        15
## 7 znakpitanja     1
```

Let's look at it by dialect:

ekavica:

```
## [1] 0.5185185
```

ijekavica:

```
## [1] 0.5813953
```

SRB:

```
## [1] 0.4782609
```

HR:

```
## [1] 0.7352941
```

BIH:

```
## [1] 0.7142857
```

The masculine form reaches threshold for Croatia. The neuter form reaches threshold for BIH. No form reaches threshold for Serbia. I don't think we should include this term considering how much gender variation there is.

cap

```
## # A tibble: 6 x 2
## # Groups:   response [6]
##   response        n
##   <chr>       <int>
## 1 kacket         24
## 2 kapa           39
## 3 kapa?           2
## 4 kapa/kacket     1
## 5 sesir           2
## 6 silterica       7
```

ekavica:

```
## [1] 0.7037037
```

ijekavica:

```
## [1] 0.6976744
```

SRB:

```
## [1] 0.7391304
```

HR:

```
## [1] 0.7352941
```

BIH:

```
## [1] 0.7142857
```

Overall this seems to be dialectal: Croatia and BIH say “kapa” and Serbia says “kacket”. Ekavica says “kacket” and Ijekavica says “kapa”. However, as a native speaker of Serbian ekavica, I consider “kacket” (cap) to be a subset of “kapa” (hat). I think we should not include this word, based on the discussion I had with Brandon discussed above.

table

```
## # A tibble: 10 x 2
## # Groups:   response [10]
##    response                    n
##    <chr>                   <int>
##  1 drvenisto                   1
##  2 hoklica                     1
##  3 klupa                      38
##  4 ripstol?                    1
##  5 sto                         7
##  6 sto.                        1
##  7 sto(prihvatitemiodgovor)    1
##  8 stol                       18
##  9 stolic                      3
## 10 stolica                     4
```

Do not include. 38 people thought this was a bench and not a table. Perhaps find a better image of a table to include.

calculator

```
## # A tibble: 6 x 2
## # Groups:   response [6]
##   response               n
##   <chr>              <int>
## 1 digitron              34
## 2 kalkulator            32
## 3 kalkulator/digitron    2
## 4 neznam                 1
## 5 racunalo               4
## 6 racunar                2
```

If we sum up digitron and kalkulator we get:

```
##           n
## 1 0.9066667
```

We reach threshold. Both words are of the same gender, we should include it.

sword

```
## # A tibble: 15 x 2
## # Groups:   response [15]
##    response          n
##    <chr>         <int>
##  1 katana            1
##  2 mac              40
##  3 mac)              1
##  4 macnoz            1
```

```
##  5 macoz             1
##  6 macsablja         2
##  7 macvina           1
##  8 mqcc              1
##  9 noz               4
## 10 pik               1
## 11 plavasablja       1
## 12 sable             1
## 13 sablja           18
## 14 samurajskimac     1
## 15 saraf             1
```

Mac (sword) and sablja (saber) are in competition, and both can be applied to this object. They are different genders, and both can be used in a single dialect (based on my own experience), so we should not include it.

ruler

```
## # A tibble: 12 x 2
## # Groups:   response [12]
##    response     n
##    <chr>    <int>
##  1 crta         1
##  2 daska        1
##  3 lenjir      30
##  4 metar        2
##  5 mjerilo      2
##  6 mjrilo       1
##  7 neznam       1
##  8 ravnala      1
##  9 ravnalo     32
## 10 redalica     1
## 11 ruler        2
## 12 vladar       1
```

We have two competing words: ravnalo (neuter) and lenjir (masculine). Let's see if it is dialectal:

ekavica:

```
## [1] 0.7777778
```

ijekavica:

```
## [1] 0.6744186
```

SRB:

```
## [1] 0.826087
```

HR:

```
## [1] 0.8235294
```

BIH:

```
## [1] 0.7142857
```

Ekavica has "lenjir" reach threshold. Ijekavica has nothing reach threshold. This is explained by the country distribution:

SRB: Lenjir reaches threshold (ekavica)
HR: ravnalo reaches threshold (ijekavica)
BIH: lenjir reaches threshold (ijekavica)

It is clear that this word is dialectal. Since I assume most of our participants will be Serbian Ekavica speakers (given community outreach), I think we should include this word.

clock

```
## # A tibble: 11 x 2
## # Groups:   response [11]
##    response           n
##    <chr>          <int>
##  1 alarm              2
##  2 alarm-sat          2
##  3 budilica          21
##  4 budilica/sat       1
##  5 budilnik          14
##  6 casovnik           2
##  7 cekerica           1
##  8 jedansat           1
##  9 sat               29
## 10 satbudilnik        1
## 11 satilicasovnik     1
```

It seems like a split that is important is: budilnik versus budilica. Let's see if those two are dialectal. If so we could include this, if not then we skip it.

ekavica:

```
## # A tibble: 6 x 2
## # Groups:   response [6]
##   response           n
##   <chr>          <int>
## 1 alarm-sat          2
## 2 budilica           1
## 3 budilnik          12
## 4 casovnik           1
## 5 cekerica           1
## 6 satilicasovnik     1
```

```
## [1] 0.6666667
```

ijekavica:

```
## [1] 0.75
```

SRB:

```
## [1] 0.7333333
```

HR:

```
## [1] 0.8636364
```

BIH:

```
## [1] 0.1666667
```

Overall it seems clear: everyone says “sat”, HR say “budilica”, SRB say “budilnik”, and for BIH there is not enough data (1 instance of budilica and 2 of budilnik). Again, because I expect we will have majority Serbian participants, I think we should include this word.

dustpan

```
## # A tibble: 36 x 2
## # Groups:   response [36]
##    response         n
##    <chr>        <int>
##  1 asov             1
##  2 cistanje         1
##  3 djubravnik       2
##  4 djubrovnik       6
##  5 drzaczametlu     1
##  6 kanta            1
##  7 kotao            1
##  8 lapota           1
##  9 levak            2
## 10 lopata           4
## # ... with 26 more rows
```

This is such a mess, we shouldn't include it.

Color Scheme 2

The number of responses >= 70% threshold in color scheme 2 split up by gender:

```
## # A tibble: 2 x 2
##   gender `n()`
##   <fct>  <int>
## 1 F         13
## 2 M         14
```

Feminine

Here are all the items in color scheme two that are feminine

```
##        item colorScheme gender      label   percent
## 1   dresser           2      F     komoda 0.4000000
## 2     ladle           2      F    kutlaca 0.4666667
## 3     purse           2      F      torba 0.4666667
## 4   whistle           2      F zvizdaljka 0.4666667
## 5    bonbon           2      F     bombon 0.4933333
## 6    basket           2      F      korpa 0.5200000
## 7    camera           2      F fotoaparat 0.5333333
## 8      fork           2      F   viljuska 0.5333333
## 9     spoon           2      F     kasika 0.5333333
## 10     duck           2      F      patka 0.5733333
## 11     bowl           2      F     zdjela 0.5866667
## 12     boot           2      F      cizma 0.6666667
## 13     vase           2      F       vaza 0.6800000
## 14   shovel           2      F     lopata 0.7333333
## 15  slipper           2      F     papuca 0.7466667
```

```
## 16 necklace           2      F    ogrlica 0.7866667
## 17   ribbon           2      F      masna 0.8266667
## 18    shell           2      F    skoljka 0.8533333
## 19     sock           2      F     carapa 0.8533333
## 20    glove           2      F   rukavica 0.8933333
## 21      tie           2      F    kravata 0.9066667
## 22      box           2      F     kutija 0.9200000
## 23     cake           2      F      torta 0.9466667
## 24     fish           2      F       riba 0.9600000
## 25    bench           2      F      klupa 0.9866667
## 26      die           2      F      kocka 1.0000000
```

As stated above, there are 13 feminine nouns that fall into category (1)

Fall above the 70% threshold
- die
- bench
- fish
- cake
- box
- tie
- glove
- sock
- shell
- ribbon
- necklace
- slipper
- shovel

Words that fall above threshold after accounting for spelling errors: - purse: (majority other responses are tasna)

Words that fall above the threshold if we combine the top words with the same gender:
- vase: {vazna, vaza} - boot: {cipela} (shoe) and {cizma} (boot) - bowl: {zdjela, cinija} - spoon: {kasika, zlica} - fork: {vilica, viljuska} - basket: {kosara, korpa} - whistle: {zvizdaljka, pistaljka}

Words that we should include contingent on dialect:
- bonbon: {bonbona} (SRB/BIH) or {bonbon} (HR)

words to consider due to compounds: - duck

Words we should not include:
- camera - ladle - dresser

Overall we should include: all words above threshold, all words that are of the same gender when combined fall above threshold
This gives us 21 target words

If we also want to include words contingent on dialect: this gives us: 22 target words

If we also want to include words with compounds: this gives us: 23 target words

Considerations for each words that fell below the 70% threshold are given below

Vase

```
## # A tibble: 13 x 2
## # Groups:   response [13]
##    response            n
```

```
##    <chr>           <int>
##  1 cup                5
##  2 cup(opet3slova)    1
##  3 glina              1
##  4 saksija            1
##  5 tegla              1
##  6 urna               3
##  7 vasa               3
##  8 vasna              1
##  9 vasnica            1
## 10 vaya               1
## 11 vaza              51
## 12 vazna              4
## 13 zelenavaza         2
```

combine: vaza and vazna

```
##           n
## 1 0.8133333
```

Reaches threshold when we combine different spellings of "vaza" and "vazna". We should include it.

Boot

```
## # A tibble: 5 x 2
## # Groups:   response [5]
##   response            n
##   <chr>           <int>
## 1 bakandza            1
## 2 cipela             22
## 3 cipelagojzerica     1
## 4 cizma              50
## 5 smedjecizma         1
```

If we combind cipela and cizma:

```
##      n
## 1 0.96
```

We get an overwhelming majority. Cipela (shoe) is a superset of cizma (boot), but they are both of the same gender so it doesn't matter. We should include them.

Bowl

```
## # A tibble: 10 x 2
## # Groups:   response [10]
##    response         n
##    <chr>        <int>
##  1 adijela          1
##  2 casa             1
##  3 cinija          19
##  4 denis            1
##  5 krigla           1
##  6 posuda           4
##  7 tanir            1
##  8 tanjir           2
##  9 zdjela          44
```

```
## 10 zelenazdjela     1
```

Combine zdela and cinija:

```
##           n
## 1 0.8533333
```

We have a majority. Zdjela and Cinija are dialectal and mean the same thing, and are of the same gender. We should include it.

Duck

```
## # A tibble: 13 x 2
## # Groups:   response [13]
##    response                   n
##    <chr>                  <int>
##  1 gumenaguska                1
##  2 gumenapatak                1
##  3 gumenapatka               17
##  4 gumenopace                 1
##  5 guska                      1
##  6 igracka                    3
##  7 ljubicastagumenapatka      1
##  8 natka                      1
##  9 patka                     43
## 10 patkaigracka               2
## 11 patkazakupanje             2
## 12 patkazakupanje/igracka     1
## 13 patkicagumena              1
```

Compounds make up:

```
##      n
## 1 0.32
```

Compounds make up for 32% of the responses, and "duck" makes up X% of the responses. If we want compounds include this one, if we really don't want compounds, exclude it.

Spoon

```
## # A tibble: 7 x 2
## # Groups:   response [7]
##   response            n
##   <chr>           <int>
## 1 kasika             40
## 2 kaska               1
## 3 ksiku               1
## 4 ljubicastazlica     1
## 5 seflja              1
## 6 zaimaca             1
## 7 zlica              30
```

```
##           n
## 1 0.9733333
```

"zlica" and "kasika" make up an overwhelming majority. Both are of the same gender and are determined by dialect, so we should include it.

Fork

```
## # A tibble: 8 x 2
## # Groups:   response [8]
##   response     n
##   <chr>    <int>
## 1 perun        1
## 2 pinjur       3
## 3 pirun        1
## 4 vilica      27
## 5 viljosku     1
## 6 viljuska    40
## 7 viluska      1
## 8 zilca        1
```

If we combine "vilica" and "viljuska":

```
##      n
## 1 0.92
```

"vilica" and "viljuska" make up an overwhelming majority. Both are of the same gender and are determined by dialect, so we should include it.

Camera

```
## # A tibble: 9 x 2
## # Groups:   response [9]
##   response              n
##   <chr>             <int>
## 1 aparatzaslikanje      2
## 2 fotic                 1
## 3 fotoaparat           40
## 4 instamatik            1
## 5 kamera               25
## 6 kamera,fotoaparat     1
## 7 kamera?               1
## 8 polaroid              3
## 9 telefon               1
```

The overwhelming responses are: fotoaparat (or some version of that) and kamera. They are of two different genders, so lets see if it is contingent on dialect:

SRB:

```
## [1] 0.6086957
```

HR:

```
## [1] 0.5588235
```

BIH:

```
## [1] 0.5
```

Fotoaparat only comes out at threshold for HR speakers. However, from personal knowledge, kamera and fotoaparat can both be used in my dialect of BCS. Therefore participants might have both at their disposal. Therfore we should not include these items.

Basket

```
## # A tibble: 7 x 2
## # Groups:   response [7]
##   response         n
##   <chr>        <int>
## 1 ceger            1
## 2 korpa           39
## 3 korpa/kosara     1
## 4 kosara          31
## 5 neznam           1
## 6 torba            1
## 7 zelenakosara     1
```

Let's combine the top two: korpa and kosara

```
##      n
## 1 0.96
```

We get an overwhelming majority. They both mean the same thing. Since both are the same gender, we should include them.

Bonbon

```
## # A tibble: 5 x 2
## # Groups:   response [5]
##   response                   n
##   <chr>                  <int>
## 1 bombon                    37
## 2 bombon#                    1
## 3 bombona                   35
## 4 ljubicastiomotzabombon     1
## 5 slatko                     1
```

The question is between bonbon (M) and bonbona (F). Is this dialectal?

Ekavica::

```
## [1] 0.8518519
```

Ijekavica::

```
## [1] 0.6976744
```

SRB:

```
## [1] 0.9130435
```

HR:

```
## [1] 0.9117647
```

BIH:

```
## [1] 0.7857143
```

It seems clear that this is dialectal: SRB: bombona HR: bombon BIH: bombona

We should include it.

whistle

```
## # A tibble: 17 x 2
## # Groups:   response [17]
##    response           n
##    <chr>          <int>
##  1 crnazvizdalica     1
##  2 frula              1
##  3 fucka              1
##  4 fuckaljka          1
##  5 pisca              1
##  6 piska              1
##  7 pistaljka         20
##  8 sviraljka          1
##  9 vesalica           1
## 10 zvidaljka          1
## 11 zvizdac            1
## 12 zvizdalica         2
## 13 zvizdaljka        35
## 14 zvizdalka          1
## 15 zvizdati           1
## 16 zvizduk            5
## 17 zvizgac            1
```

```
##           n
## 1 0.7866667
```

Pistaljka and Zvizdaljka together give the threshold and both are feminine. We shoudl include it.

purse

```
## # A tibble: 9 x 2
## # Groups:   response [9]
##   response              n
##   <chr>             <int>
## 1 ljubicastatorbica     1
## 2 novcanik              1
## 3 rucnatorba            1
## 4 tasna                16
## 5 tasnu                 1
## 6 torba                35
## 7 torbe                 1
## 8 torbica              18
## 9 torbiva               1
```

Does torba/torbica make it on its own:

```
##           n
## 1 0.7466667
```

Torba on its own meets the threshold. The competitor is "tasna" which is also feminine. So I think we shoudl include this one.

ladle

```
## # A tibble: 20 x 2
## # Groups:   response [20]
```

```
##    response                   n
##    <chr>                  <int>
##  1 grabilica                  1
##  2 grabilicazajuhu            1
##  3 kaciola                    4
##  4 kacivola                   1
##  5 kasika                     1
##  6 kaska                      1
##  7 kuhaca                     2
##  8 kuhinjskavelikakasika      1
##  9 kutlaca                   35
## 10 kutljaca                   6
## 11 ladle                      1
## 12 neznam                     2
## 13 neznam,onozavaditijuhu     1
## 14 paljaca                    1
## 15 sefarka                    1
## 16 sefla                      1
## 17 seflja                     9
## 18 susak                      1
## 19 velikazlica                1
## 20 zaimaca                    4
```

This one is so over the place that I think we should not include it.

dresser

```
## # A tibble: 20 x 2
## # Groups:   response [20]
##    response              n
##    <chr>             <int>
##  1 fijoke                1
##  2 fioka                 1
##  3 garderoba             2
##  4 kabinet               1
##  5 komoda               30
##  6 kredenac              3
##  7 ladicar               4
##  8 ladice                2
##  9 latice                1
## 10 natkasna              2
## 11 neznam                1
## 12 nocniormar            1
## 13 odelastolnjak         1
## 14 ormar                19
## 15 ormaricsaladicama     1
## 16 ormarladicar          1
## 17 polica                1
## 18 regal                 1
## 19 stalaza               1
## 20 zeleniormar           1
```

It comes down to komoda and ormar. Both are of different genders, and both exist in my dialect which means that participants can choose between the two. Therefore, I think we should exclude this one.

Masculine

Here are all the items in color scheme two that are masculine

```
##           item colorScheme gender      label   percent
## 1         rope           2      M    konopac 0.2266667
## 2  screwdriver           2      M  srafciger 0.2800000
## 3       sponge           2      M     spuzva 0.3600000
## 4     lipstick           2      M     karmin 0.4133333
## 5         lock           2      M    katanac 0.4533333
## 6        razor           2      M     brijac 0.4933333
## 7       remote           2      M  daljinski 0.5200000
## 8   binoculars           2      M    dvogled 0.5866667
## 9        plate           2      M     tanjir 0.6133333
## 10       radio           2      M      radio 0.6666667
## 11       knife           2      M        noz 0.6933333
## 12     bandaid           2      M    flaster 0.7333333
## 13     luggage           2      M      kofer 0.7733333
## 14       robot           2      M      robot 0.8400000
## 15      shield           2      M       stit 0.8400000
## 16  helicopter           2      M helikopter 0.8800000
## 17  microscope           2      M  mikroskop 0.8800000
## 18        drum           2      M     bubanj 0.8933333
## 19        tent           2      M      sator 0.9066667
## 20       piano           2      M     klavir 0.9200000
## 21       truck           2      M     kamion 0.9200000
## 22      hammer           2      M      cekic 0.9466667
## 23    airplane           2      M      avion 0.9733333
## 24    umbrella           2      M   kisobran 0.9733333
## 25    calendar           2      M   kalendar 0.9866667
```

As stated above, there are 14 masculine nouns that fall into category (1)

Fall above the 70% threshold
- calendar
- umbrella
- airplane
- hammer
- truck
- piano
- tent
- drum
- microcope
- helicopter
- shield
- robot
- luggage
- bandaid

Words that fall above threshold after accounting for spelling errors: - knife

Words that fall above the threshold if we combine the top words with the same gender:
- radio: {radio, kasetefon}
- tanjir: {tanjir, tanjur} (! beware of “tacna” which appeared often and is feminine!!!!) - binoculars: {dvogled, dalekozor} - lock: {lokot, katanac} - lipstick: {karmin, ruz} - screwdriver: {srafciger, odvijac}

Words that we should include contingent on dialect:

- razor: {brijac} (SRB/BIH), {britvica} (HR) - sponge: {sundjer} (SRB), {spuzva} (HR/BIH) - rope: {kanap} (SRB), {kanap/uze/spaga} (HR/BIH)

Words contingent on compounds: - remote (60% without compounds)

Words we should not include:

Overall we should include: all words above threshold, all words that are of the same gender when combined fall above threshold
This gives us 21 target words

If we also want to include words contingent on dialect:
this gives us: 24 target words

If we also want to include words with compounds:
this gives us: 25 target words

Considerations for each words that fell below the 70% threshold are given below

knife

```
## # A tibble: 18 x 2
## # Groups:   response [18]
##    response                     n
##    <chr>                    <int>
##  1 crninoz                      1
##  2 knife                        1
##  3 kuharskinoz                  1
##  4 kuhinjskinoz                 7
##  5 ljubicastnoz                 1
##  6 narancastinoz                1
##  7 noz                         52
##  8 noz.                         1
##  9 noz...                       1
## 10 noz(opetnecerecod3slova)     1
## 11 nozh                         1
## 12 nozz                         1
## 13 nozzarezanje                 1
## 14 nozzasecenje                 1
## 15 ovojenoz                     1
## 16 sjeckalo                     1
## 17 velikinoz                    1
## 18 zeleninoz                    1
```

```
##     n
## 1 0.8
```

"Noz" reaches threshold, include it.

radio

```
## # A tibble: 16 x 2
## # Groups:   response [16]
##    response                n
##    <chr>               <int>
##  1 boombox                 1
##  2 kaseta                  1
##  3 kasetar/radio           1
```

```
##  4 kasetofon               8
##  5 kasetofonradio          1
##  6 kazetofon               4
##  7 kazic                   1
##  8 ljubicastikasetofon     1
##  9 muzic                   1
## 10 neznam                  1
## 11 radio                  50
## 12 radion                  1
## 13 radiosakazetama         1
## 14 stereo                  1
## 15 tranzistor              1
## 16 zeleniradio             1
```

Combine "kasetefon"/"kasetar" (Casette player) and radio (radio):

```
##           n
## 1 0.8666667
```

By compining "kasetofon" (cassette player) and radio (radio) we reach majority. I think we shoudl include it.

plate

```
## # A tibble: 13 x 2
## # Groups:   response [13]
##    response               n
##    <chr>              <int>
##  1 cepzakaduililavabo     1
##  2 dugma                  1
##  3 dugme                  1
##  4 frizbi                 1
##  5 narancastitanjir       1
##  6 pijat                  1
##  7 tacna                  9
##  8 tanir                  1
##  9 tanjir                46
## 10 tanjiric               4
## 11 tanjuric               7
## 12 tanjuriczasalicu       1
## 13 tasnica                1
```

Account for spelling errors and "tanjur"/"tanjir" dialect difference:

```
##     n
## 1 0.8
```

This reaches threshold. Keep in mind that the 9 other responses for "tacna" which is feminine, applies to this object, and is found in the dialect which has "tanjir". We should include this word.

binoculars

```
## # A tibble: 6 x 2
## # Groups:   response [6]
##   response      n
##   <chr>     <int>
## 1 binokular     1
## 2 blank         1
```

```
## 3 dalekozor    24
## 4 dvogled      44
## 5 kanocal       1
## 6 neznam        4
```

If we combine dalekozor and dvogled:

```
##           n
## 1 0.9066667
```

We reach threshold. Both are masculine. We should include it.

remote

```
## # A tibble: 20 x 2
## # Groups:   response [20]
##    response                   n
##    <chr>                  <int>
##  1 dalijniski                 1
##  2 dalinski                   1
##  3 dalinskiupravac            1
##  4 daljenski                  1
##  5 daljinski                 39
##  6 daljinskiupavljac          1
##  7 daljinskiupravljac        17
##  8 daljinskiupravljaczatv     1
##  9 daljinskiuprevljac         1
## 10 daljinskiuprvaljac         1
## 11 daljisnki                  1
## 12 dalkjinski                 1
## 13 daniljski                  1
## 14 kontrola                   1
## 15 mjenjaczatelevizor         1
## 16 neznam                     1
## 17 remota                     1
## 18 remote                     1
## 19 remotecontrol              1
## 20 upravljac                  2
```

Spelling errors:

```
##     n
## 1 0.6
```

Lets add in the compound:

```
##           n
## 1 0.8666667
```

We reach threshold if we include compounds (remot control versus remote). I would advice against including this compound because it is especially long.

razor

```
## # A tibble: 11 x 2
## # Groups:   response [11]
##    response            n
##    <chr>           <int>
```

```
##  1 bic                1
##  2 brijac            37
##  3 brijac/britvica    1
##  4 brijalica          1
##  5 britvica          22
##  6 neznam             1
##  7 razer              1
##  8 strugac            1
##  9 usisivac           1
## 10 zilet              8
## 11 ziletzabrijanje    1
```

It comes down to britva versus brijac: Lets look at it dialectally:

Ekavica::

```
## [1] 0.6666667
```

Ijekavica::

```
## [1] 0.4186047
```

SRB:

```
## [1] 0.7391304
```

HR:

```
## [1] 0.5294118
```

Note that HR has 8 responses for "brijac" and 18 for "britvica", 5 for "zilet"

BIH:

```
## [1] 0.7857143
```

SRB and BIH reach threshold for Brijac. HR is split betwen britvica and brijac, however accounts for the majority of cases for "britvica". I recommend including it.

lock

```
## # A tibble: 14 x 2
## # Groups:   response [14]
##    response            n
##    <chr>           <int>
##  1 brava               5
##  2 katanac            34
##  3 katanac/lokot       1
##  4 kljucanica          1
##  5 ljubicastilokot     1
##  6 lock                1
##  7 lokna               1
##  8 lokot              25
##  9 neznam              1
## 10 pramen              1
## 11 privezak            1
## 12 zakljucak           1
## 13 zaklucak            1
## 14 zaklucat            1
```

Majority items are Lokot and Katanac. Together:

```
##           n
## 1 0.8133333
```

Reaches threshold. I think we should include it since both are masculine.

lipstick

```
## # A tibble: 10 x 2
## # Groups:   response [10]
##    response            n
##    <chr>           <int>
##  1 karmin             31
##  2 karmin/ruz          2
##  3 rumenilo            1
##  4 ruz                25
##  5 ruz/karmin          1
##  6 ruzzausne           6
##  7 sminka              6
##  8 sminkaruz           1
##  9 zeleniruz           1
## 10 zeleniruzzausne     1
```

Combine karmin and ruz:

```
##     n
## 1 0.8
```

Ruz and karmin together reaches threshold. Both are masculine. We should include it.

sponge

```
## # A tibble: 17 x 2
## # Groups:   response [17]
##    response                n
##    <chr>               <int>
##  1 brisac                  1
##  2 brisanje                1
##  3 cetka                   1
##  4 cistac                  1
##  5 gumica                  1
##  6 klupa                   3
##  7 neznam                  1
##  8 pecat                   1
##  9 spuzva                 37
## 10 spuzva,skocbrajt        1
## 11 spuzvazapranjesuda      2
## 12 spuzvazapranjesudja     1
## 13 spuzvazasude            1
## 14 sudnjer                 1
## 15 sundjer                20
## 16 tablugumica             1
## 17 tabure                  1
```

Spuzva versus sundjer –> is it dialectal?

Ekavica::

```
## [1] 0.7407407
```

Ijekavica::

```
## [1] 0.6744186
```

SRB:

```
## [1] 0.826087
```

HR:

```
## [1] 0.7647059
```

BIH:

```
## [1] 0.5
```

This is clearly dialectal: SRB: Sundjer (M) HR: spuzva (F) BIH: spuzva (F) –> although it did not reach threshold, 5 pariticipants from BIH did not recognize the object.

I think we should keep it, but remember these dialect differences.

screwdriver

```
## # A tibble: 9 x 2
## # Groups:   response [9]
##   response               n
##   <chr>              <int>
## 1 izvijac                1
## 2 kacavida               4
## 3 karcavida              1
## 4 neznam                 1
## 5 odvijac               17
## 6 srafciger             48
## 7 srafcigersrafciger     1
## 8 tool                   1
## 9 veslo                  1
```

Combining odvijac and srafciger:

```
##           n
## 1 0.8666667
```

Reaches threshold. Srafciger and odvijac are dialectal, but both are masucline, so we should include it.

rope

```
## # A tibble: 11 x 2
## # Groups:   response [11]
##    response       n
##    <chr>      <int>
##  1 cabel          1
##  2 kabal          1
##  3 kanap         43
##  4 niz            1
##  5 niz/struna     1
##  6 snjura         1
##  7 spaga          9
##  8 uze           15
```

```
##  9 uzica          1
## 10 vrpce          1
## 11 vuna           1
```

It is between: uze, kanap, and uze. Let's see if they are dialectal

Ekavica::

```
## [1] 0.7777778
```

Ijekavica::

```
## [1] 0.4651163
```

SRB:

```
## [1] 0.826087
```

HR:

```
## [1] 0.3529412
```

Kanap (12), uze (11), spaga (6)

BIH:

```
## [1] 0.6428571
```

Again this seems to be dialectal. If we think we will have majority serbian participants, we should include it.

Overall

If we only include words that meet threshold:
CS1:
- Fem: 16
- Masc: 12

CS2:
- Fem: 14
- Masc: 15

If we include words with same gender terms
CS1:
- Fem: 19
- Masc: 14

CS2:
- Fem: 21
- Masc: 21

If we include dialect dependent words
CS1:
- Fem: 21
- Masc: 17

CS2:
- Fem: 22
- Masc: 24

If we include compounds
CS1:
- Fem: 22
- Masc: 19

CS2:
- Fem: 23
- Masc: 25

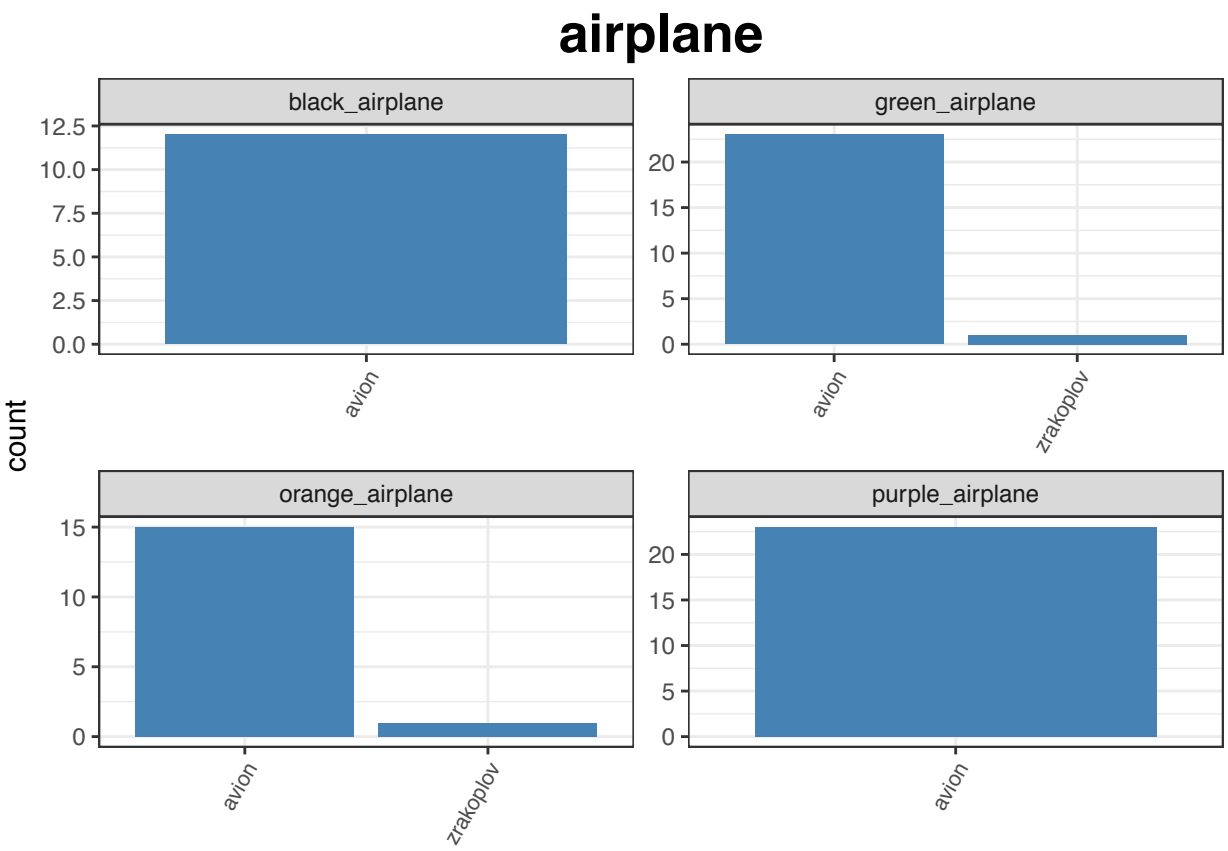
It really comes down to the main question: Do we want to include objects with multiple labels of different genders, and if so will that be compatible with our analysis?

And do we want to include compounds (billiard ball and rubber duck)?

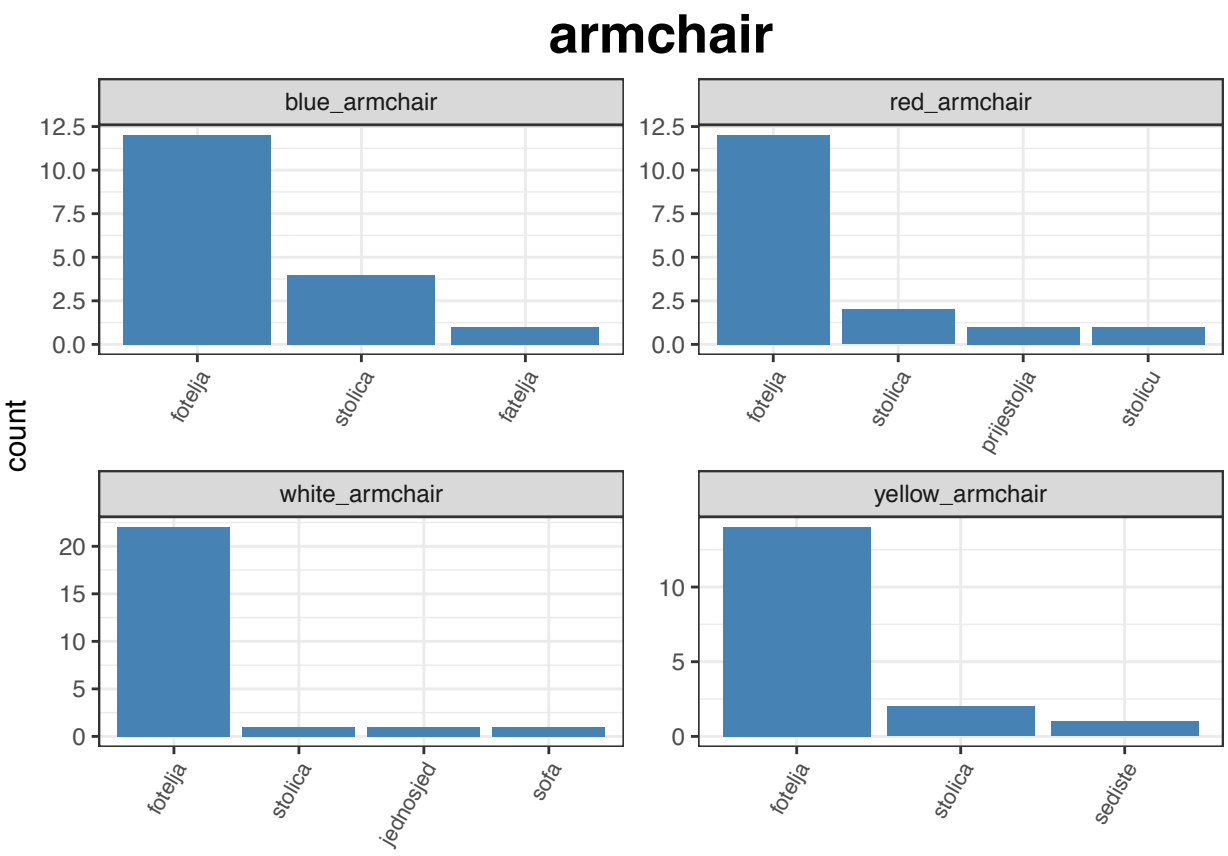
Participant Responses

If you want to see tables of the above information, uncomment the thing below:

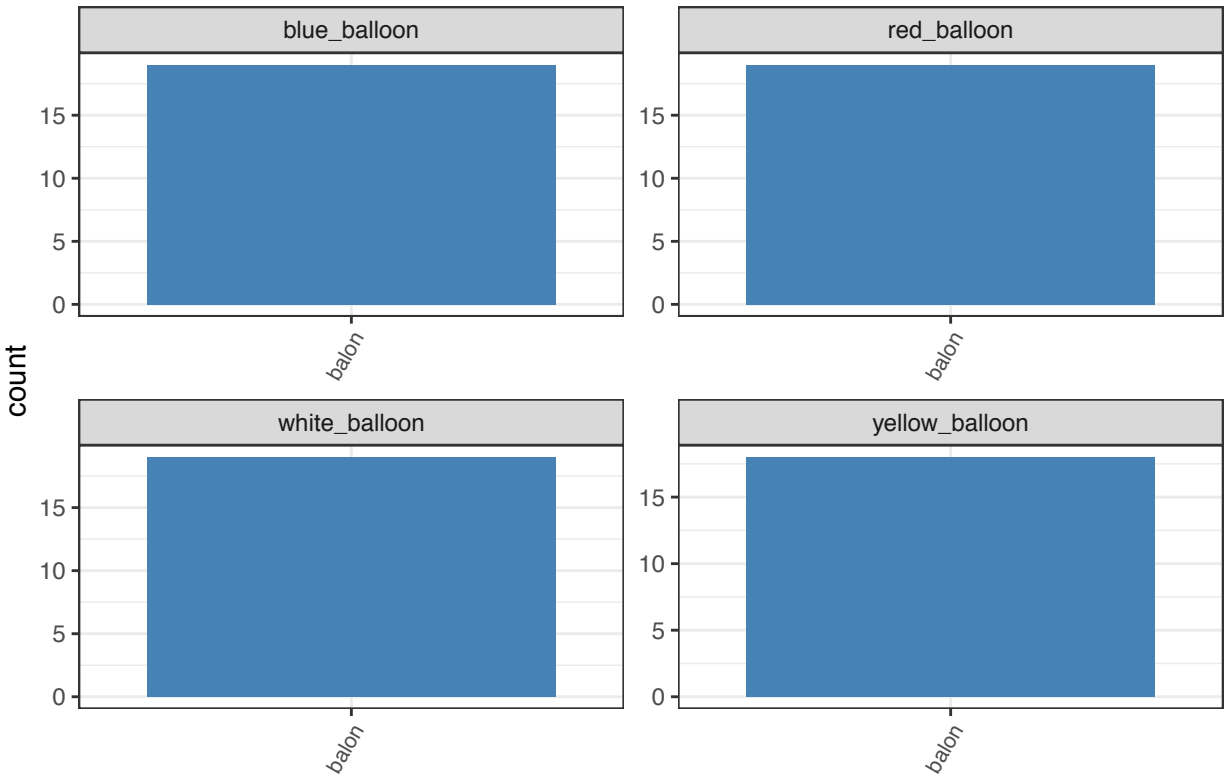
```
## [[1]]
```



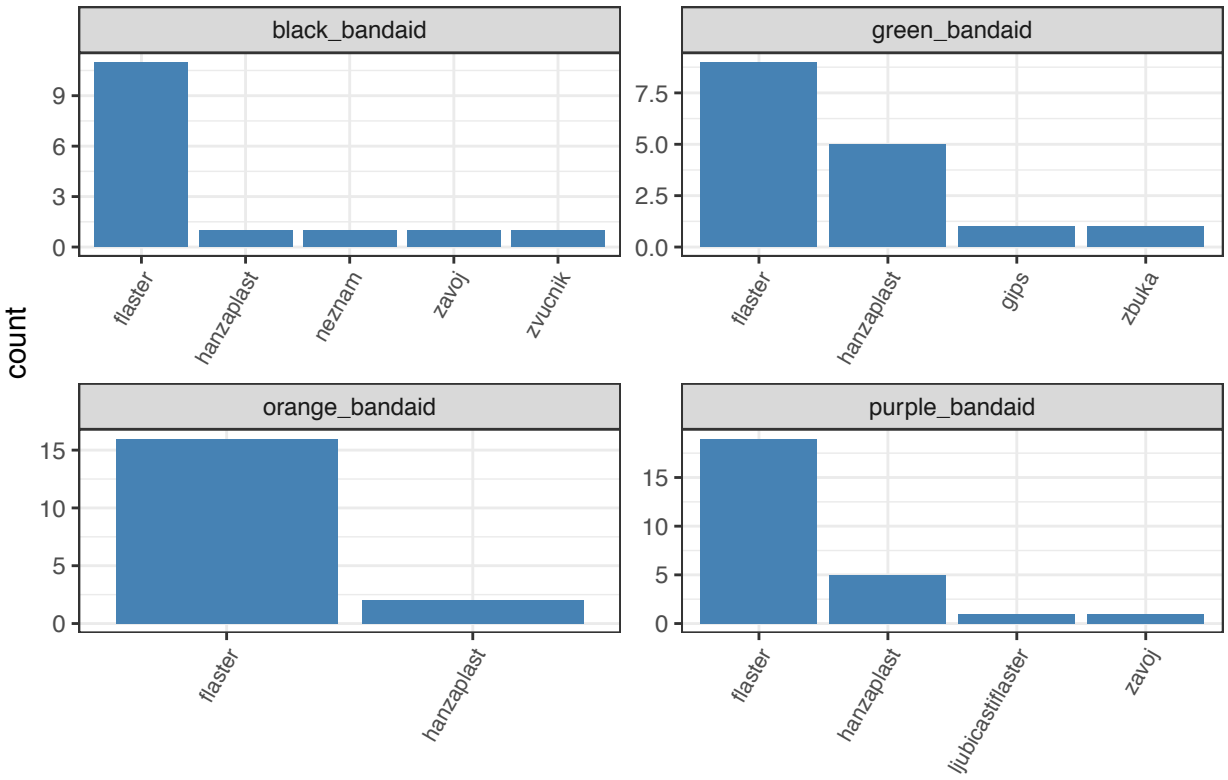

```
##
## [[2]]
```


armchair
blue_armchair
red_armchair
white_armchair
yellow_armchair
count
12.5
10.0
7.5
5.0
2.5
0.0
20
15
10
5
0
fotelja
stolica
fatelja
prijestolja
stolicu
jednosjed
sofa
sediste

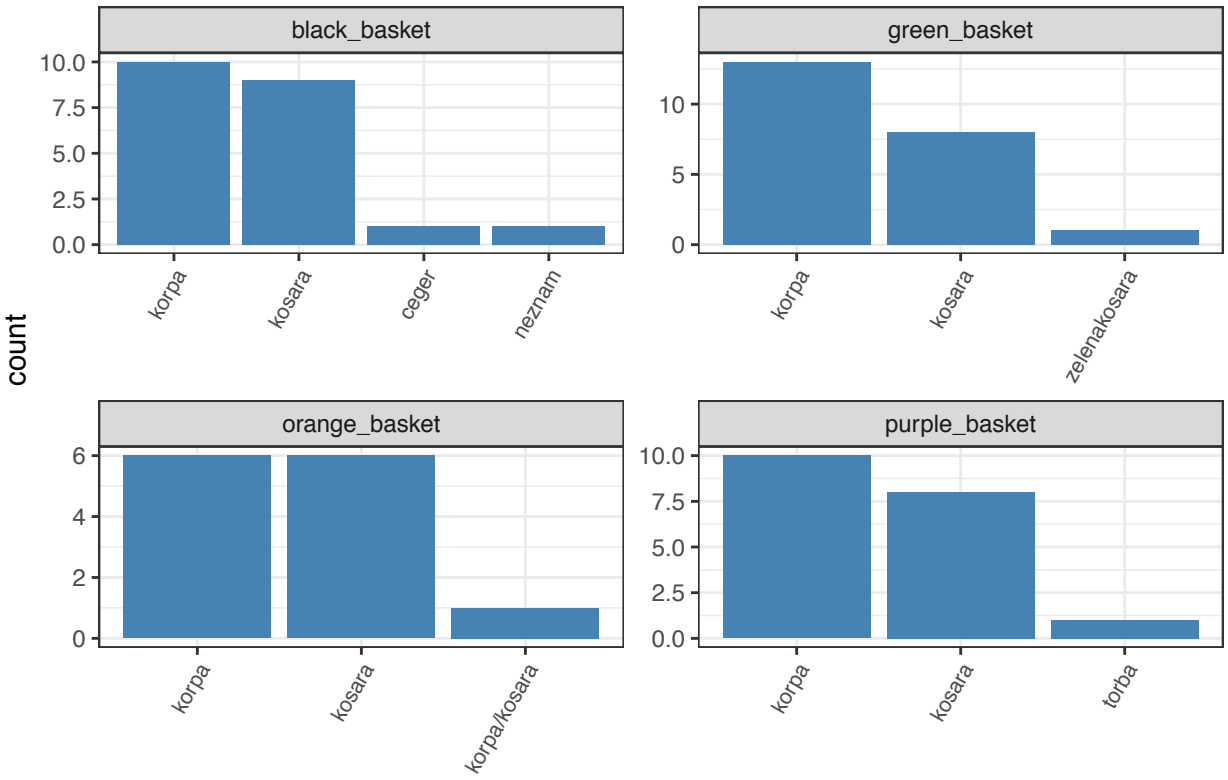
```
##
## [[3]]
```


balloon
blue_balloon
red_balloon
white_balloon
yellow_balloon
count
0
5
10
15
balon

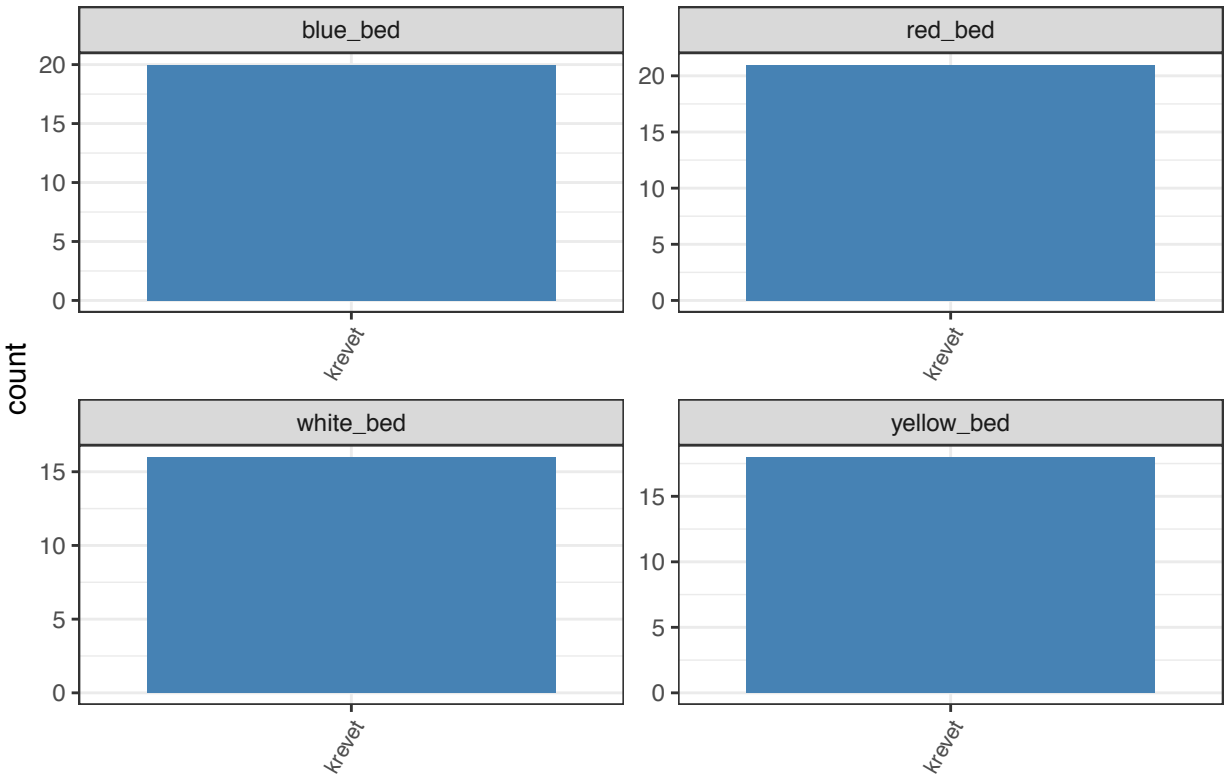
```
##
## [[4]]
```


bandaid
black_bandaid
green_bandaid
orange_bandaid
purple_bandaid
count
flaster
hanzaplast
neznam
zavoj
zvucnik
gips
zbuka
ljubicastiflaster

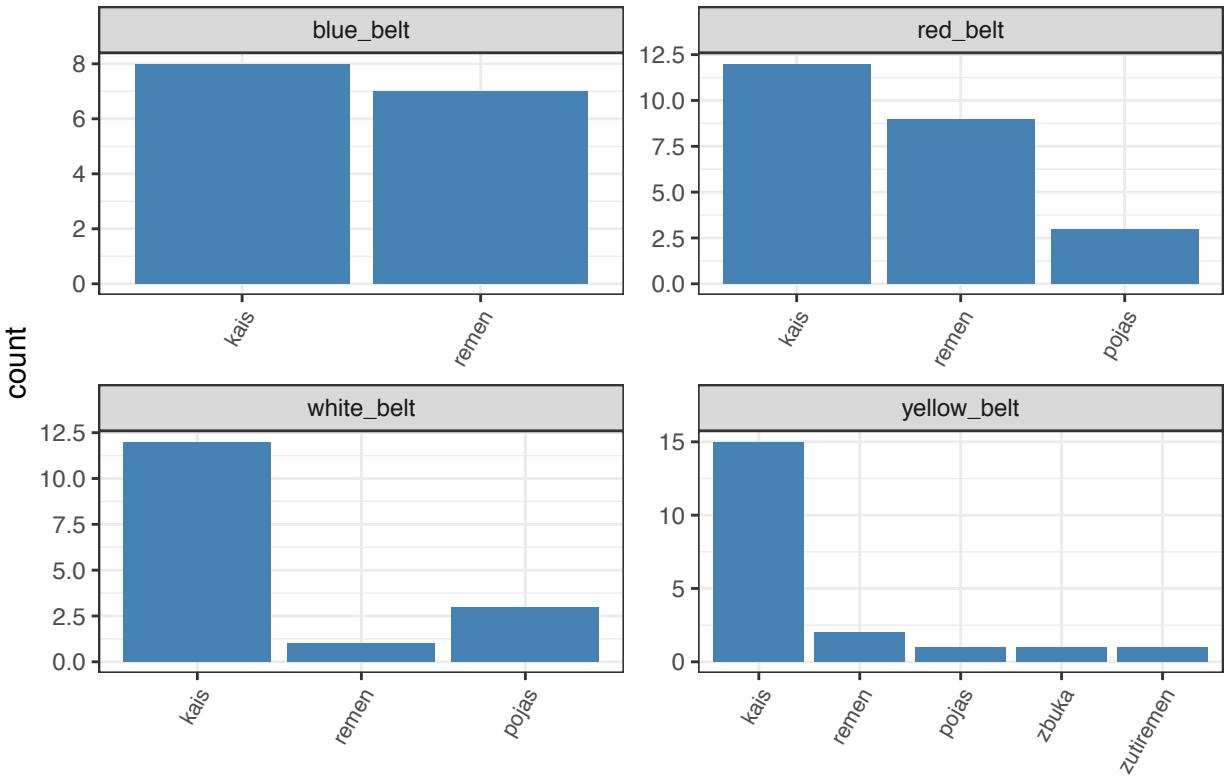
```
##
## [[5]]
```


basket
black_basket
green_basket
orange_basket
purple_basket
count
10.0
7.5
5.0
2.5
0.0
10
5
0
6
4
2
korpa
kosara
ceger
neznam
zelenakosara
korpa/kosara
torba

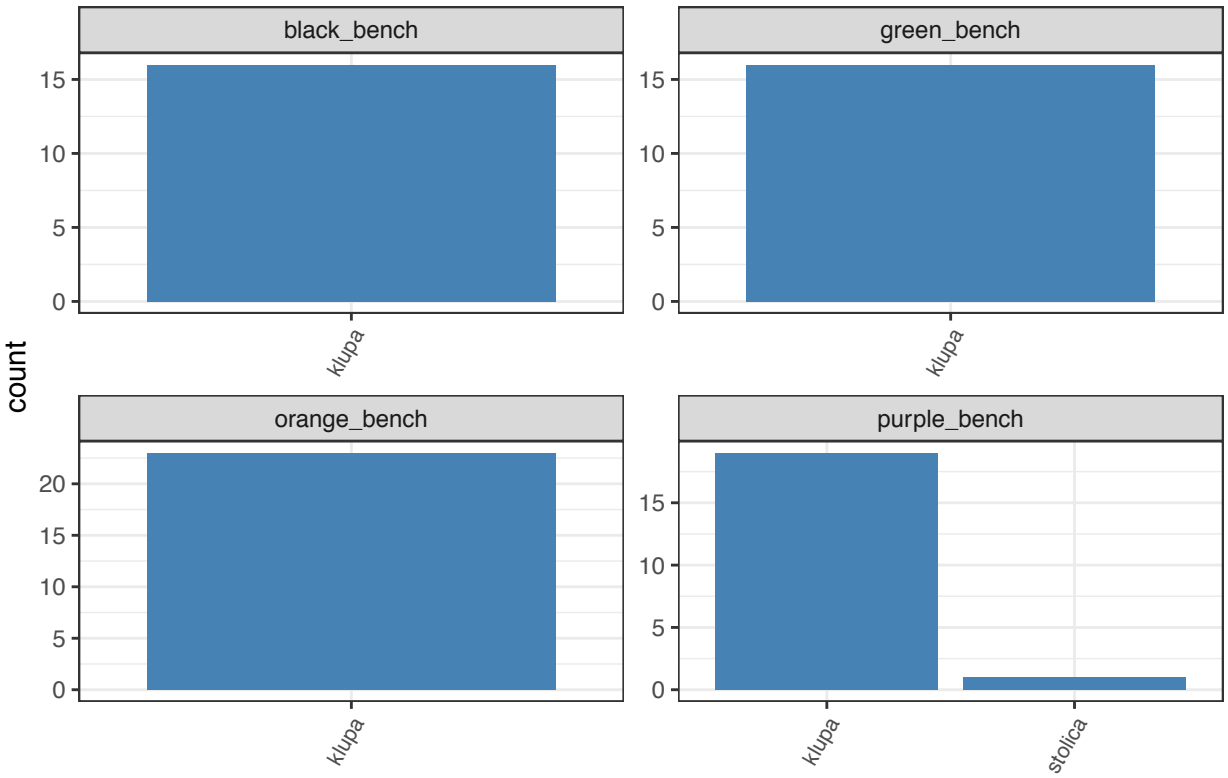
```
##
## [[6]]
```


bed
blue_bed
red_bed
white_bed
yellow_bed
count
krevet
0
5
10
15
20

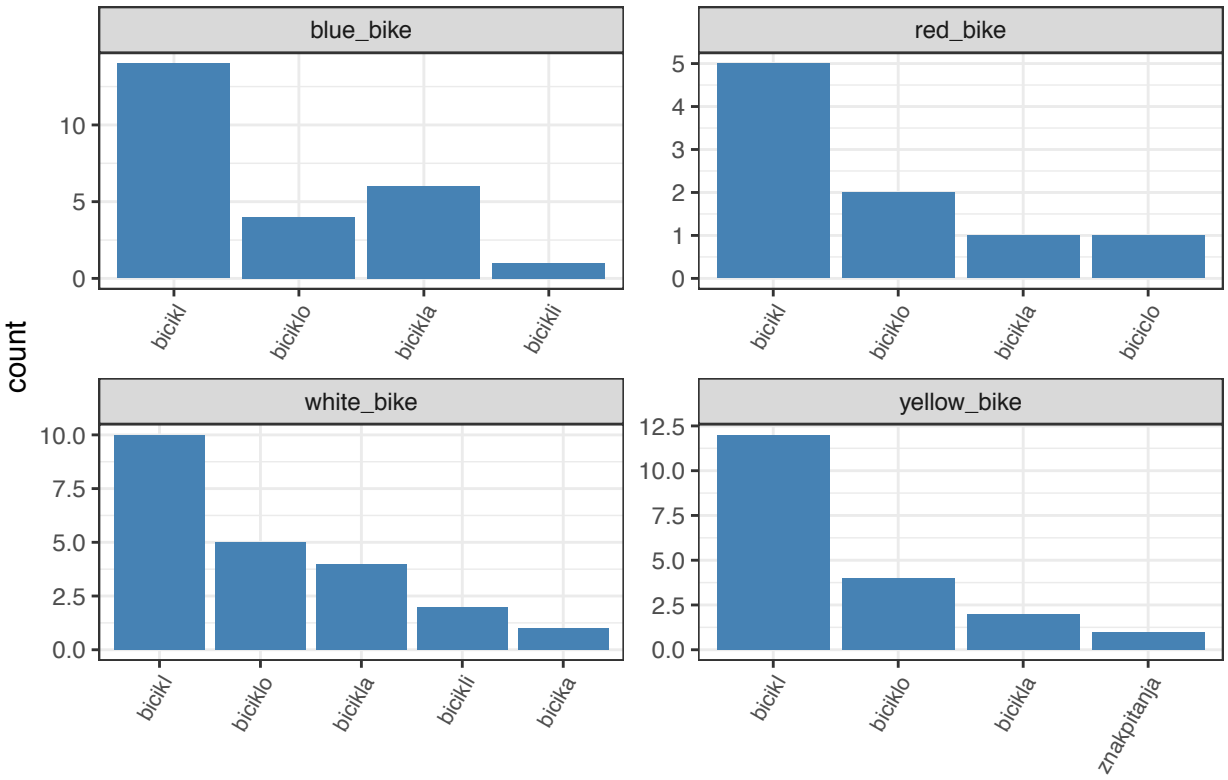
```
##
## [[7]]
```


belt
blue_belt
red_belt
white_belt
yellow_belt
count
kais
remen
pojas
zbuka
zutiremen

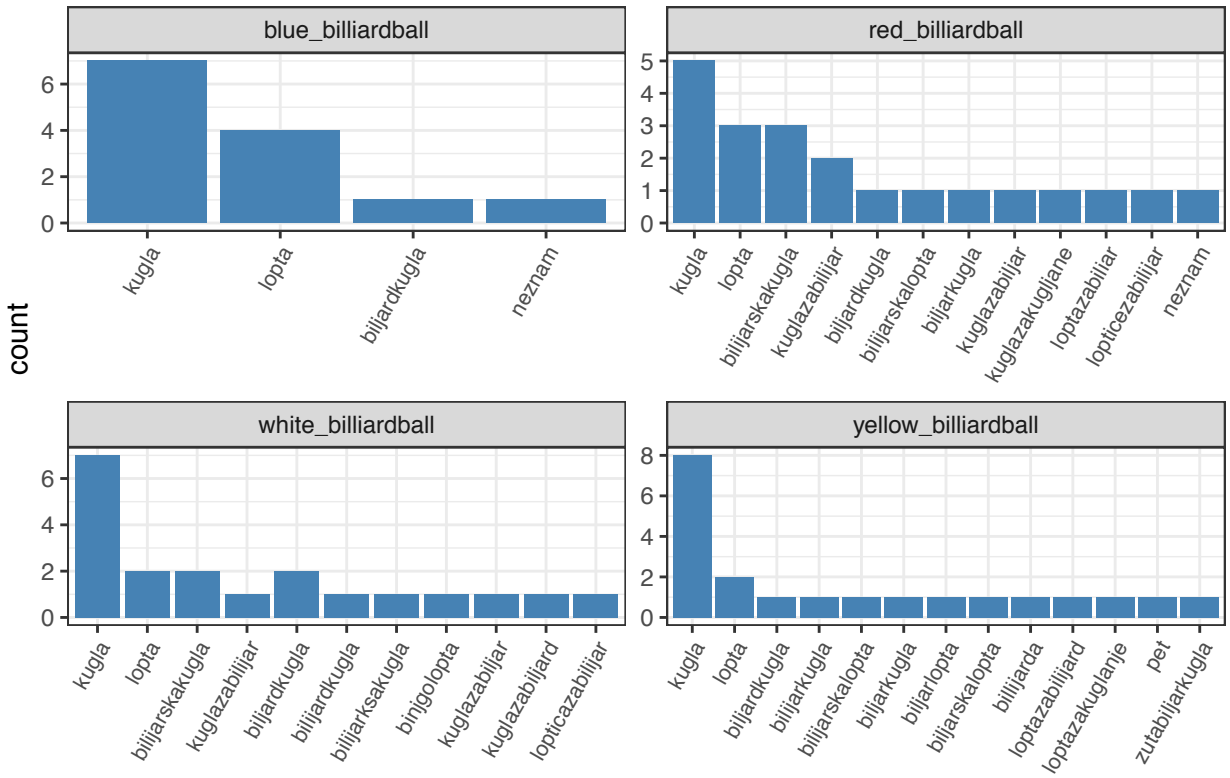
```
##
## [[8]]
```


bench
black_bench
green_bench
orange_bench
purple_bench
count
0
5
10
15
20
klupa
stolica

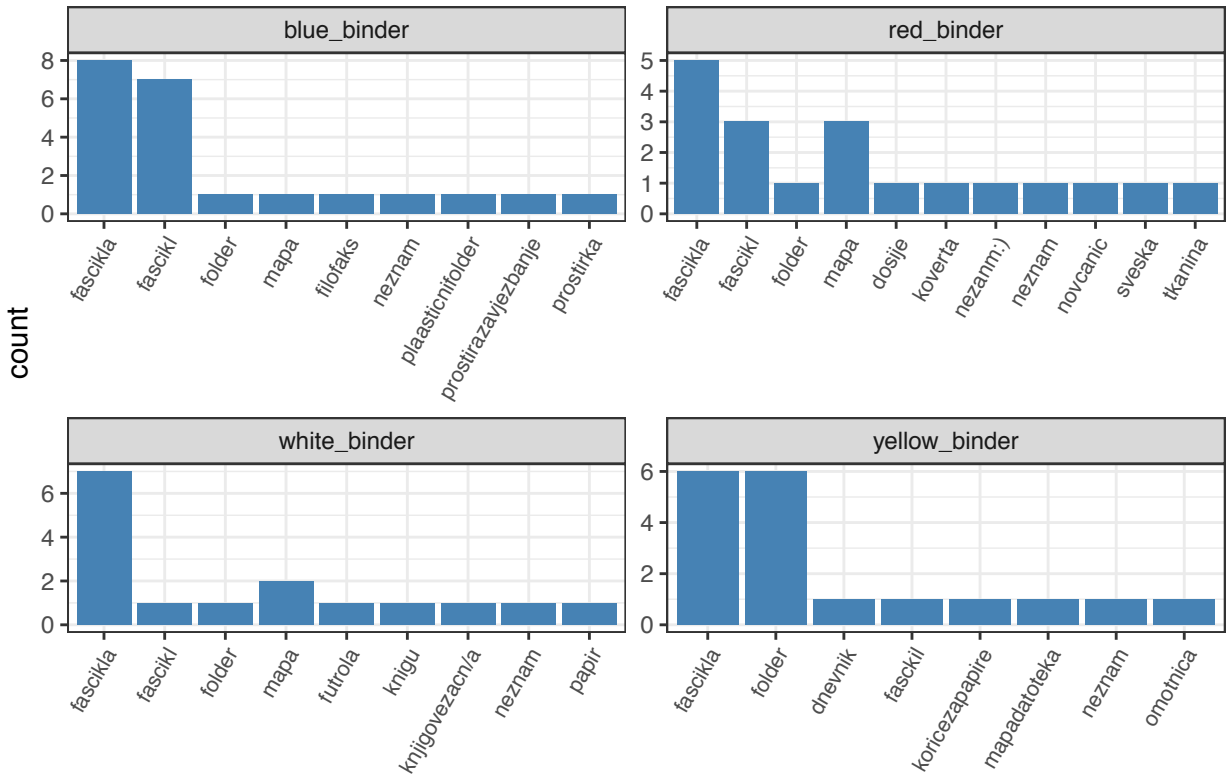
```
##
## [[9]]
```


bike
blue_bike
red_bike
white_bike
yellow_bike
count
bicikl
biciklo
bicikla
biciklı
bicika
biciclo
znakpitanja

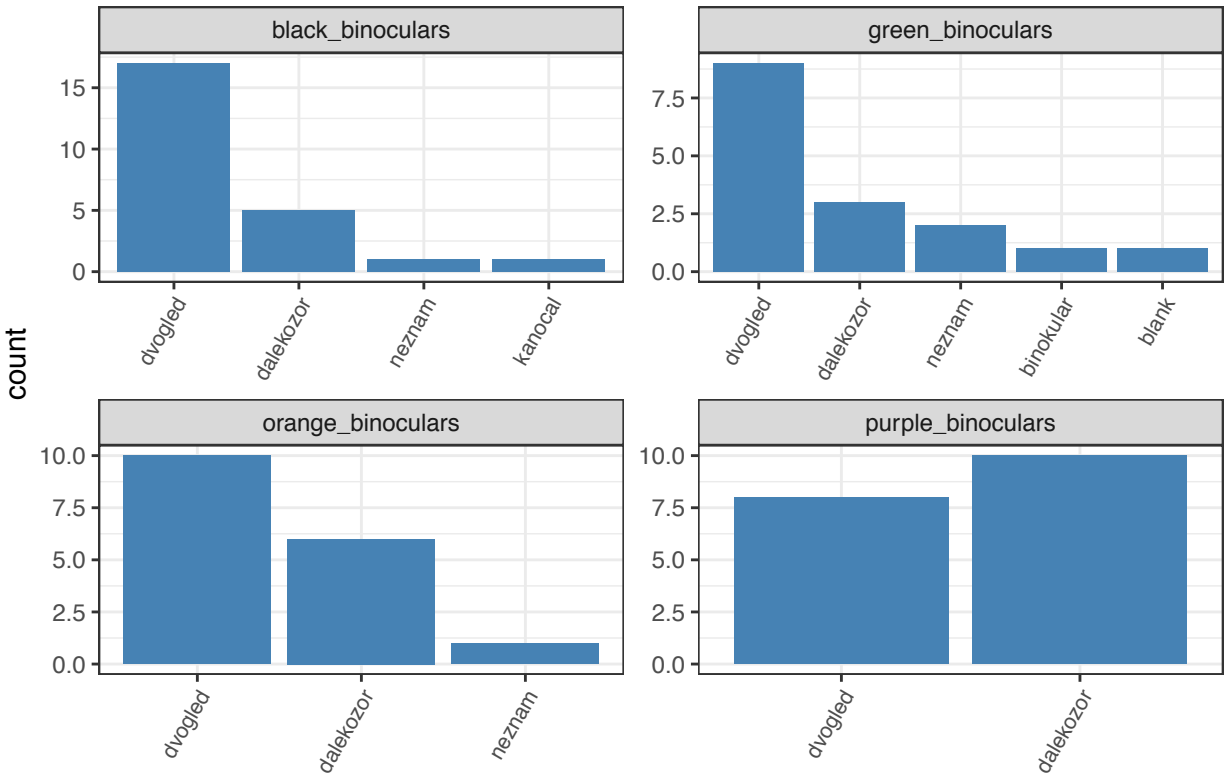
```
##
## [[10]]
```


billiardball
blue_billiardball
red_billiardball
white_billiardball
yellow_billiardball
count
kugla
lopta
biljardkugla
neznam
bilijarskakugla
kuglazabilijar
biljardkugla
bilijarskalopta
biljarkugla
kuglazakugljane
loptazabiliar
lopticezabilijar
bilijarksakugla
binjgolopta
kuglazabiljard
lopticazabilijar
biljarlopta
biljarskalopta
bilijarda
loptazabilijard
loptazakuglanje
pet
zutabiljarkugla

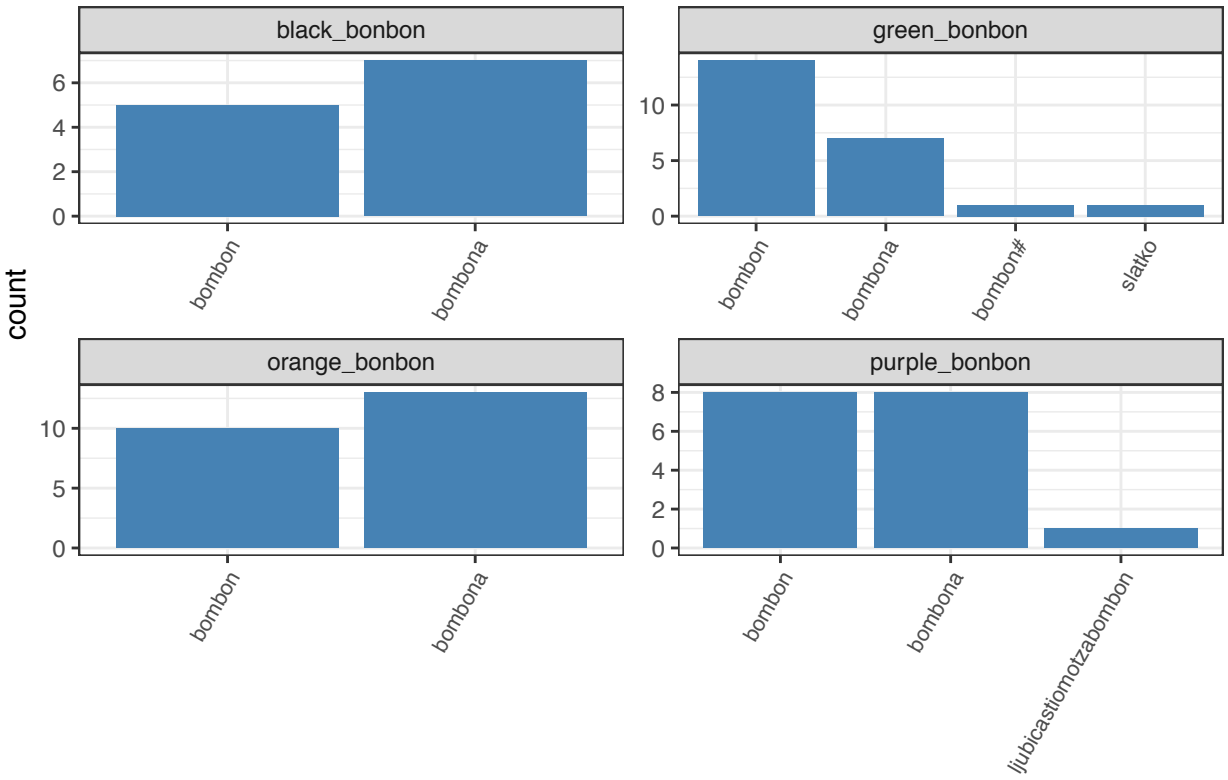
```
##
## [[11]]
```


binder
blue_binder
red_binder
white_binder
yellow_binder
count
fascikla
fascikl
folder
mapa
filofaks
neznam
plasticnifolder
prostirazavjezbanje
prostirka
dosije
koverta
nezanm:)
novcanic
sveska
tkanina
futrola
knigu
knjigovezacn/a
papir
dnevnik
fasckil
koricezapapire
mapadatoteka
omotnica

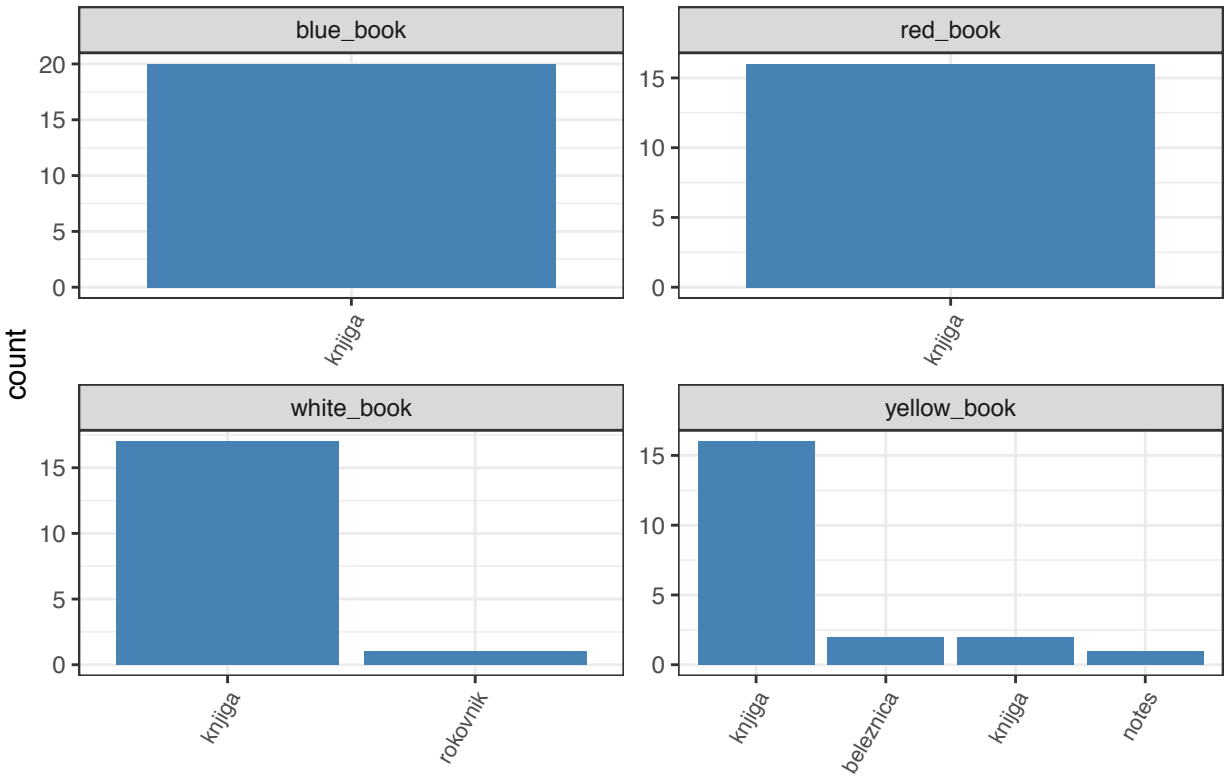
```
##
## [[12]]
```


binoculars
black_binoculars
green_binoculars
orange_binoculars
purple_binoculars
count
dvogled
dalekozor
neznam
kanocal
binokular
blank

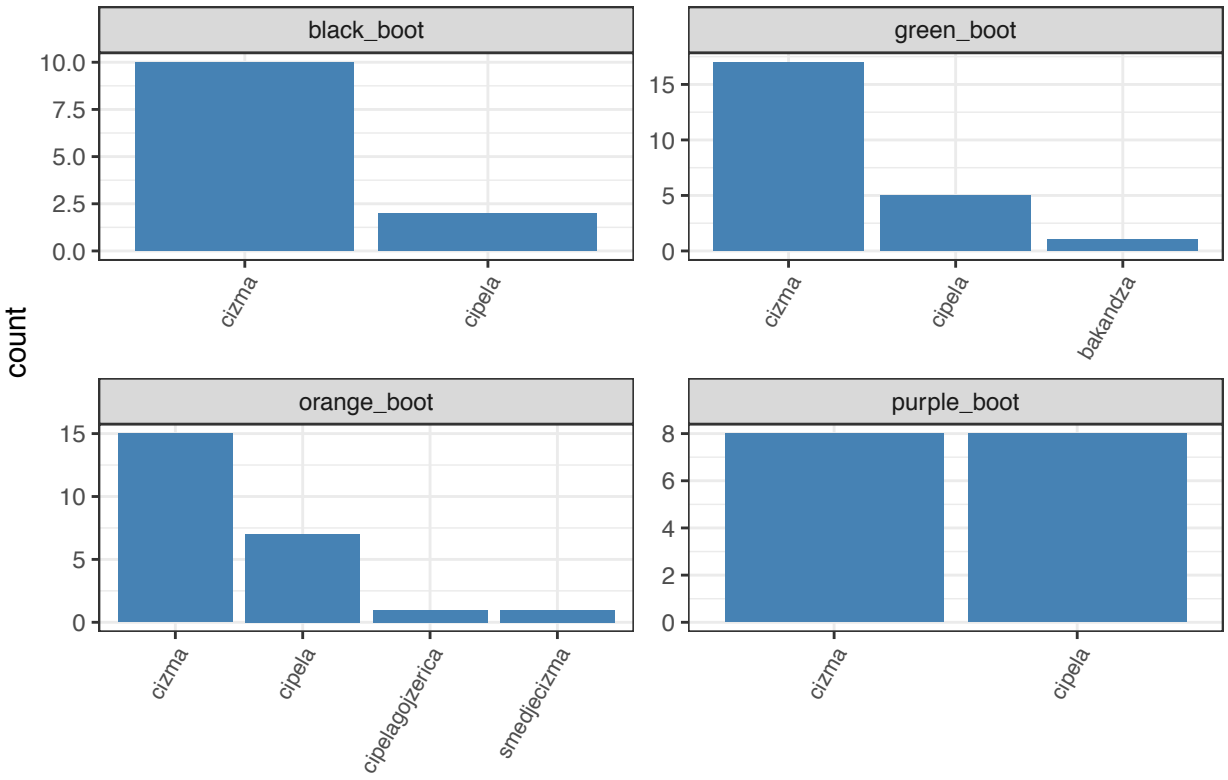
```
##
## [[13]]
```


bonbon
black_bonbon
green_bonbon
orange_bonbon
purple_bonbon
count
bombon
bombona
bombon#
slatko
ljubicastiomotzabombon

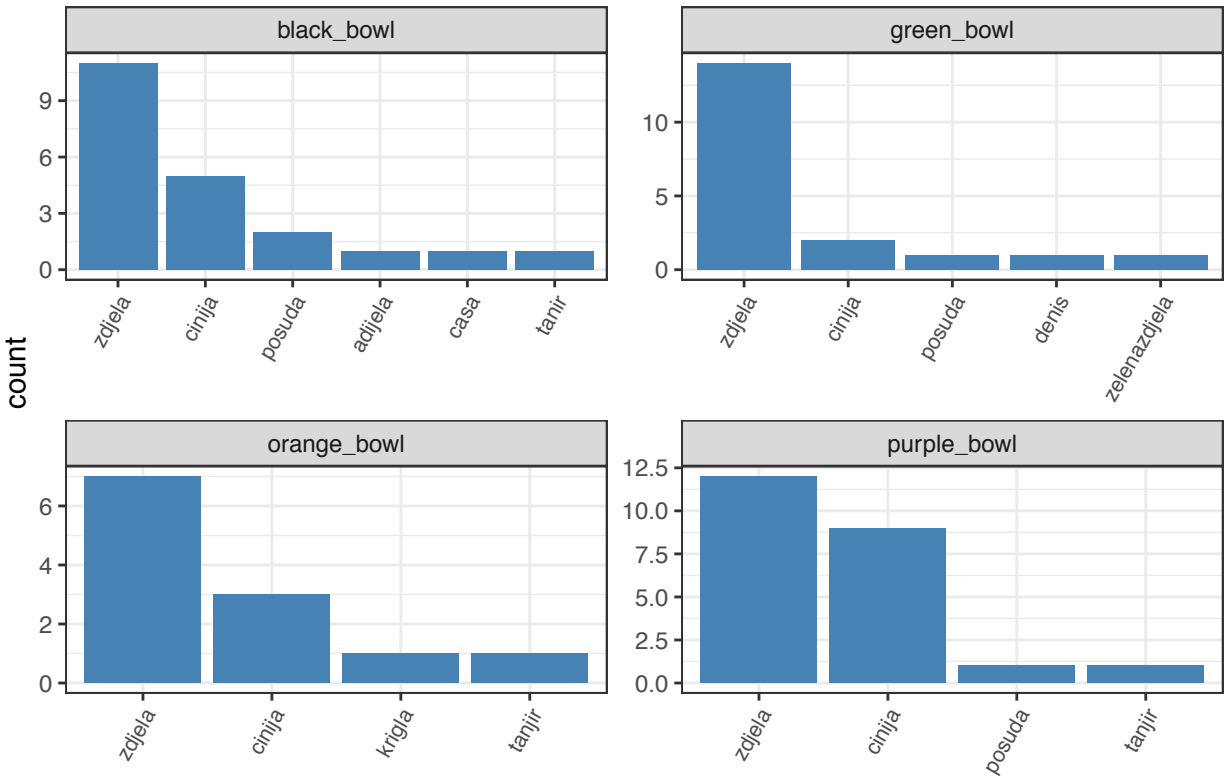
```
##
## [[14]]
```


book
blue_book
red_book
white_book
yellow_book
count
0
5
10
15
20
knjiga
rokovnik
beleznica
knilga
notes

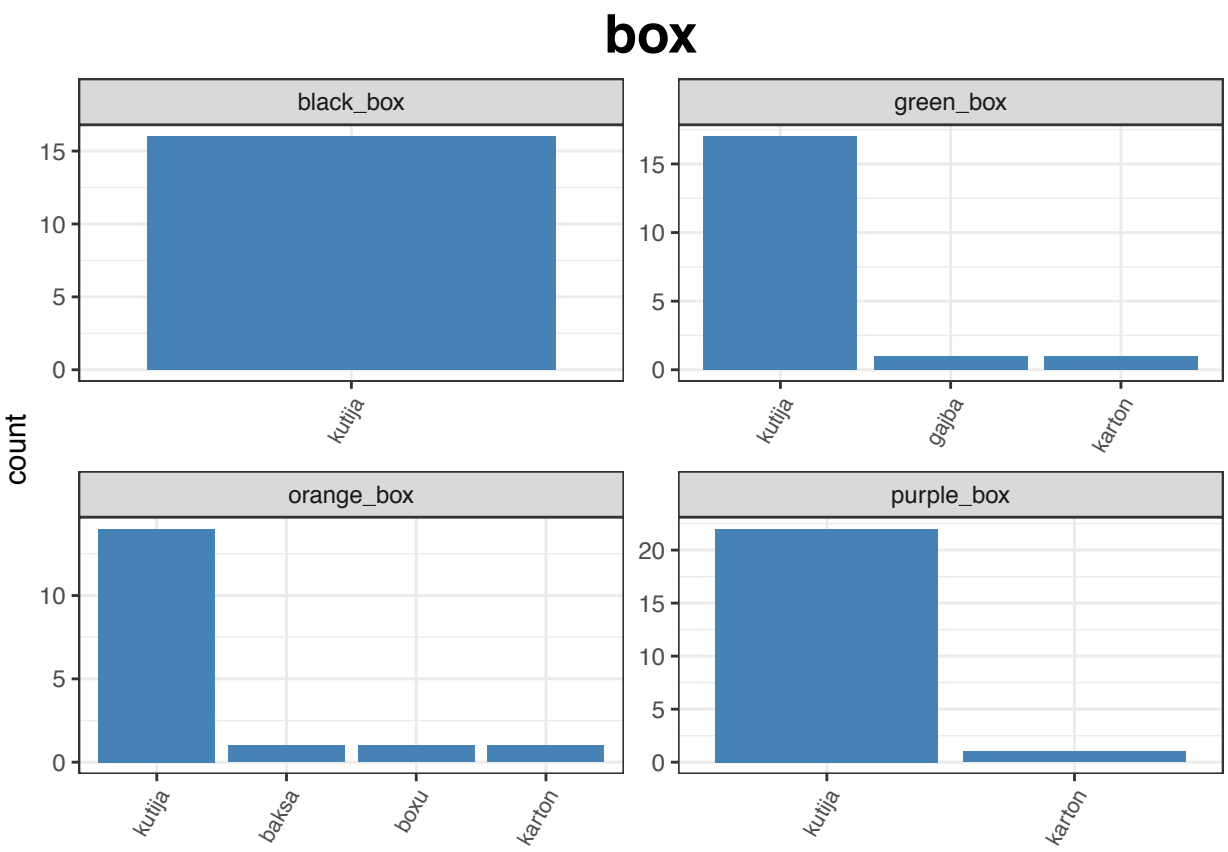
```
##
## [[15]]
```


boot
black_boot
green_boot
orange_boot
purple_boot
count
10.0
7.5
5.0
2.5
0.0
15
10
5
0
8
6
4
2
cizma
cipela
bakandza
cipelagojzerica
smedjecizma

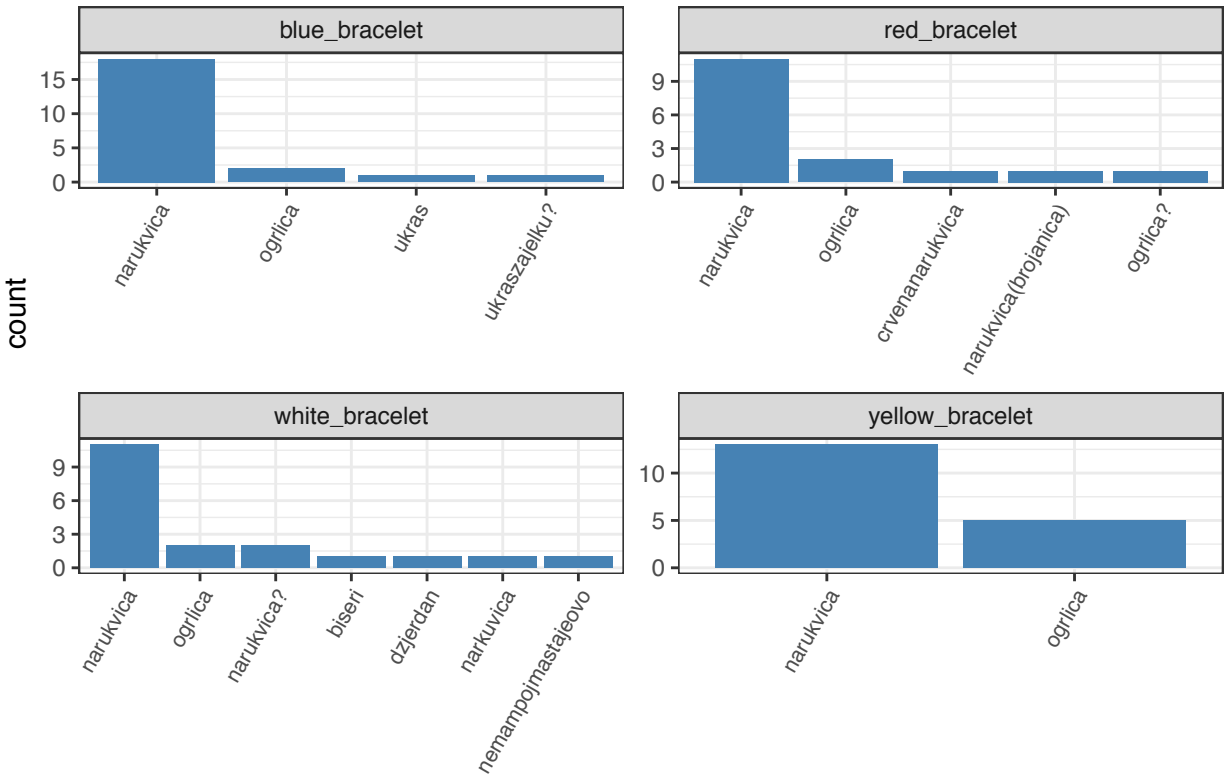
```
##
## [[16]]
```


bowl
black_bowl
green_bowl
orange_bowl
purple_bowl
count
zdjela
cinija
posuda
adijela
casa
tanir
denis
zelenazdjela
krigla
tanjir

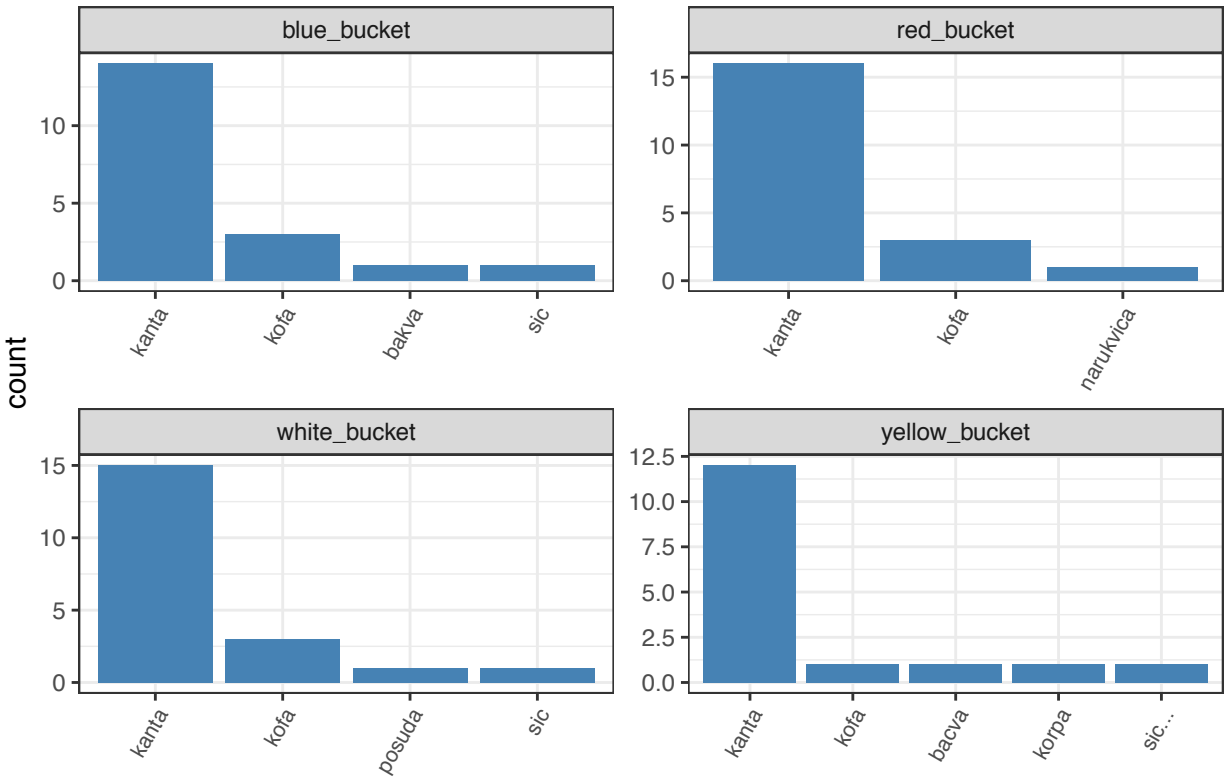
```
##
## [[17]]
```


box
black_box
green_box
orange_box
purple_box
count
kutija
gajba
karton
baksa
boxu

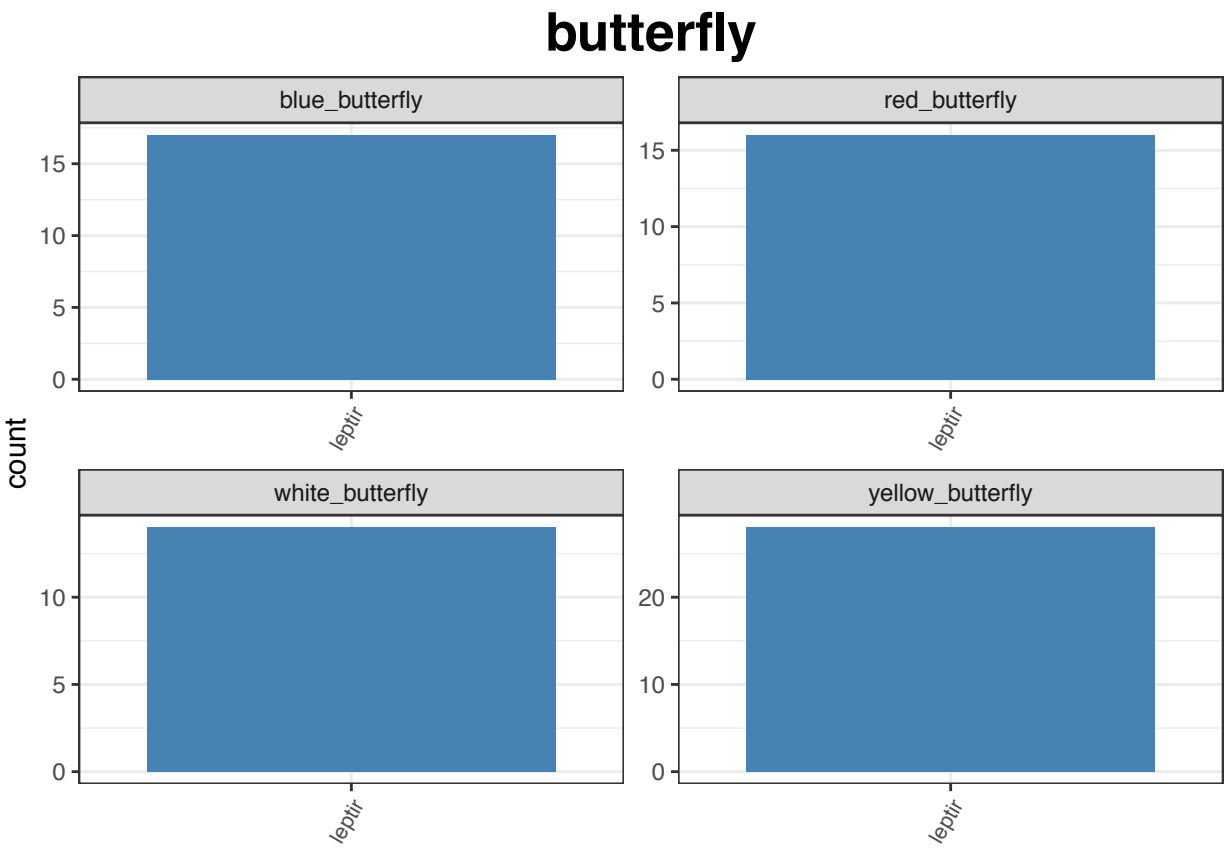
```
##
## [[18]]
```


bracelet
blue_bracelet
red_bracelet
white_bracelet
yellow_bracelet
count
narukvica
ogrlica
ukras
ukraszajelku?
crvenanarukvica
narukvica(brojanica)
ogrlica?
biseri
dzjerdan
narkuvica
nemampojmastajeovo

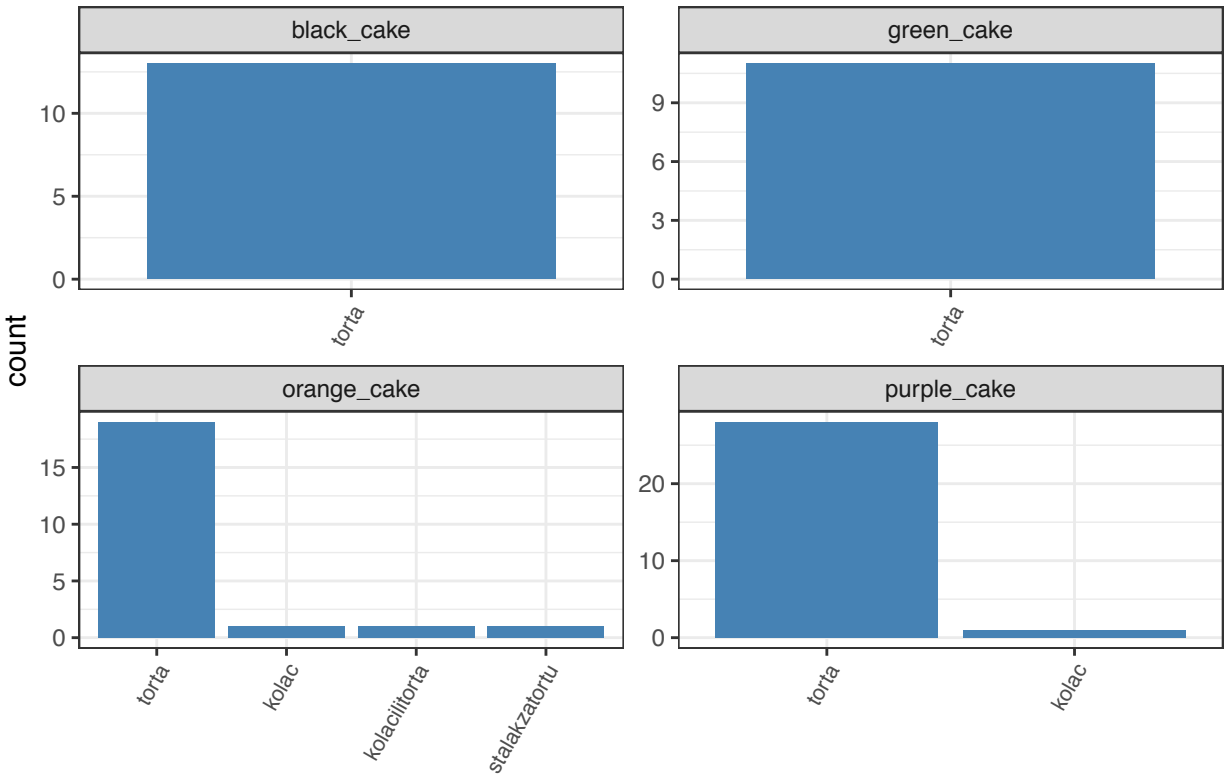
```
##
## [[19]]
```


bucket
blue_bucket
red_bucket
white_bucket
yellow_bucket
count
kanta
kofa
bakva
sic
narukvica
posuda
bacva
korpa
sic...

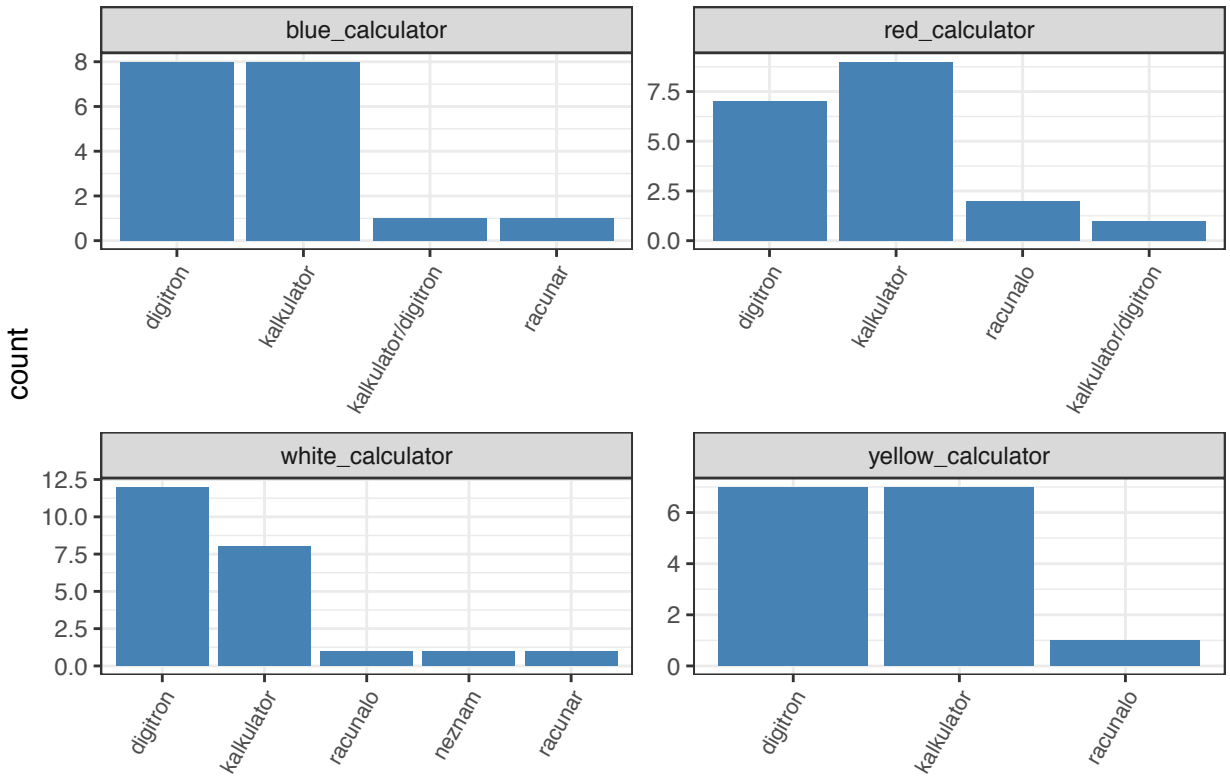
```
##
## [[20]]
```


butterfly
blue_butterfly
red_butterfly
white_butterfly
yellow_butterfly
count
leptir

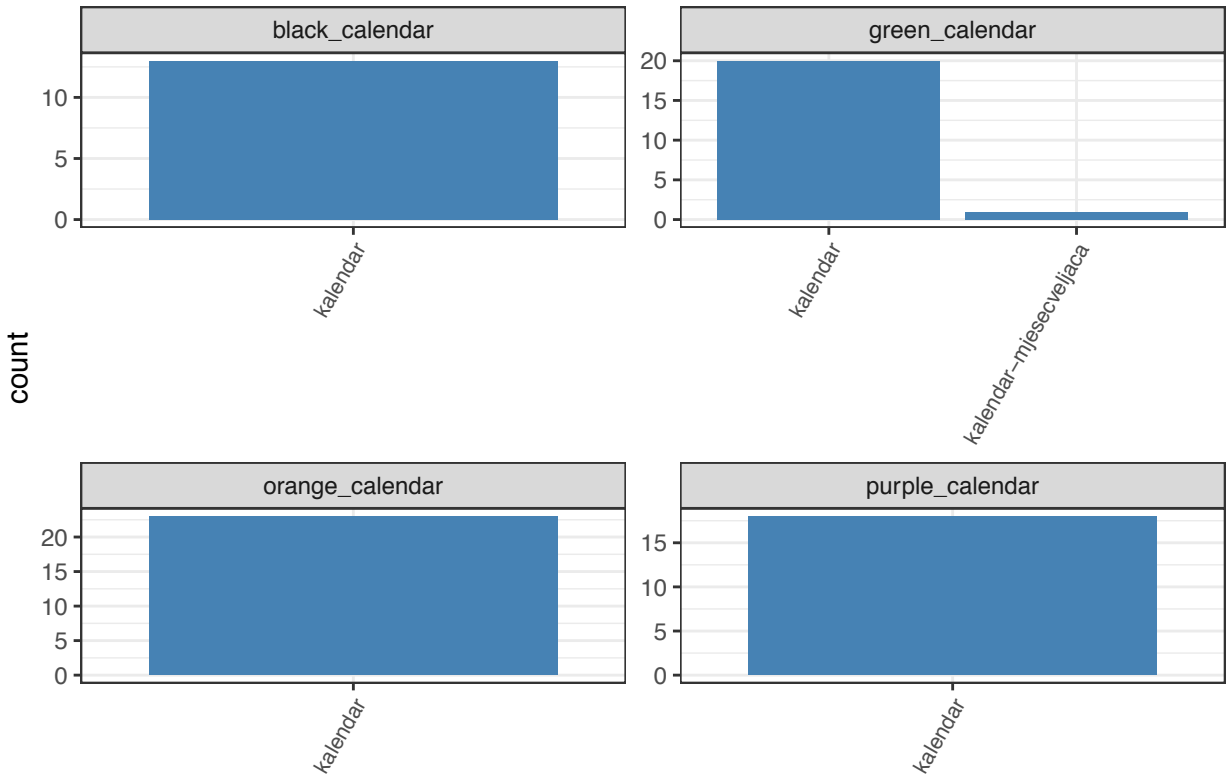
```
##
## [[21]]
```


cake
black_cake
green_cake
orange_cake
purple_cake
count
torta
kolac
kolacilitorta
stalakzatortu

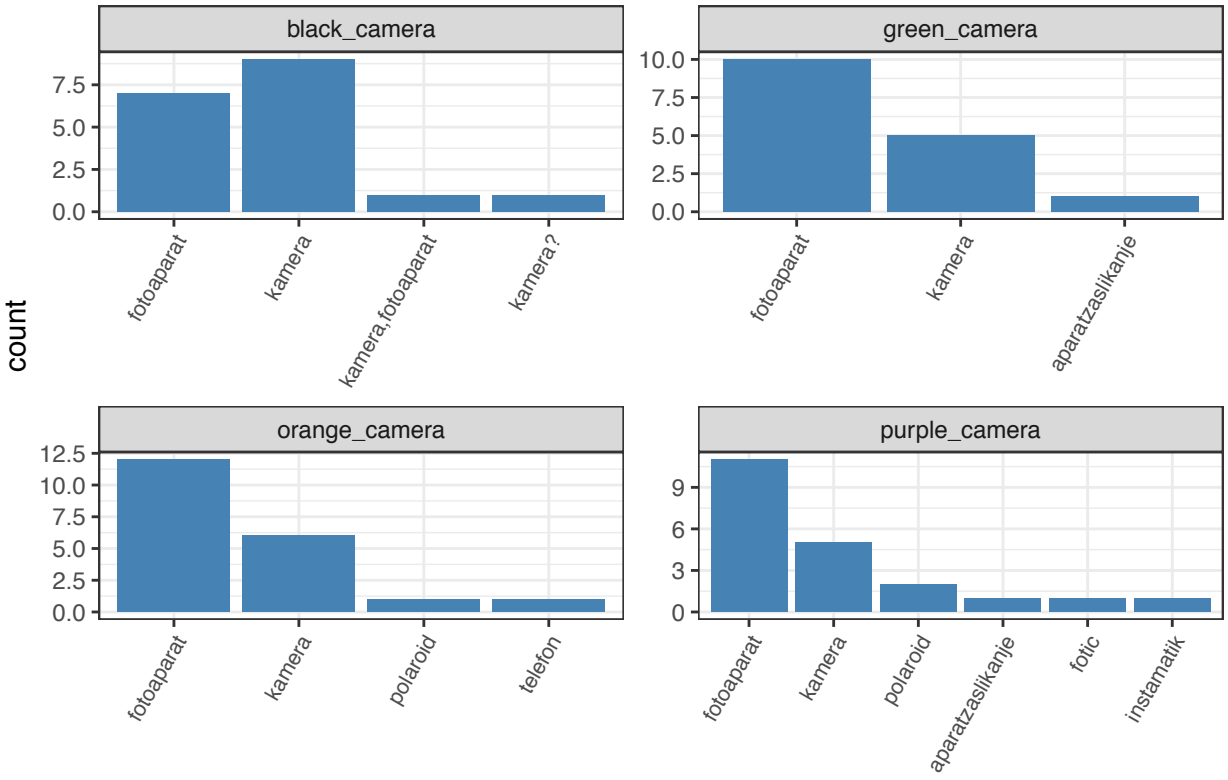
```
##
## [[22]]
```


calculator
blue_calculator
red_calculator
white_calculator
yellow_calculator
count
digitron
kalkulator
kalkulator/digitron
racunar
racunalo
neznam

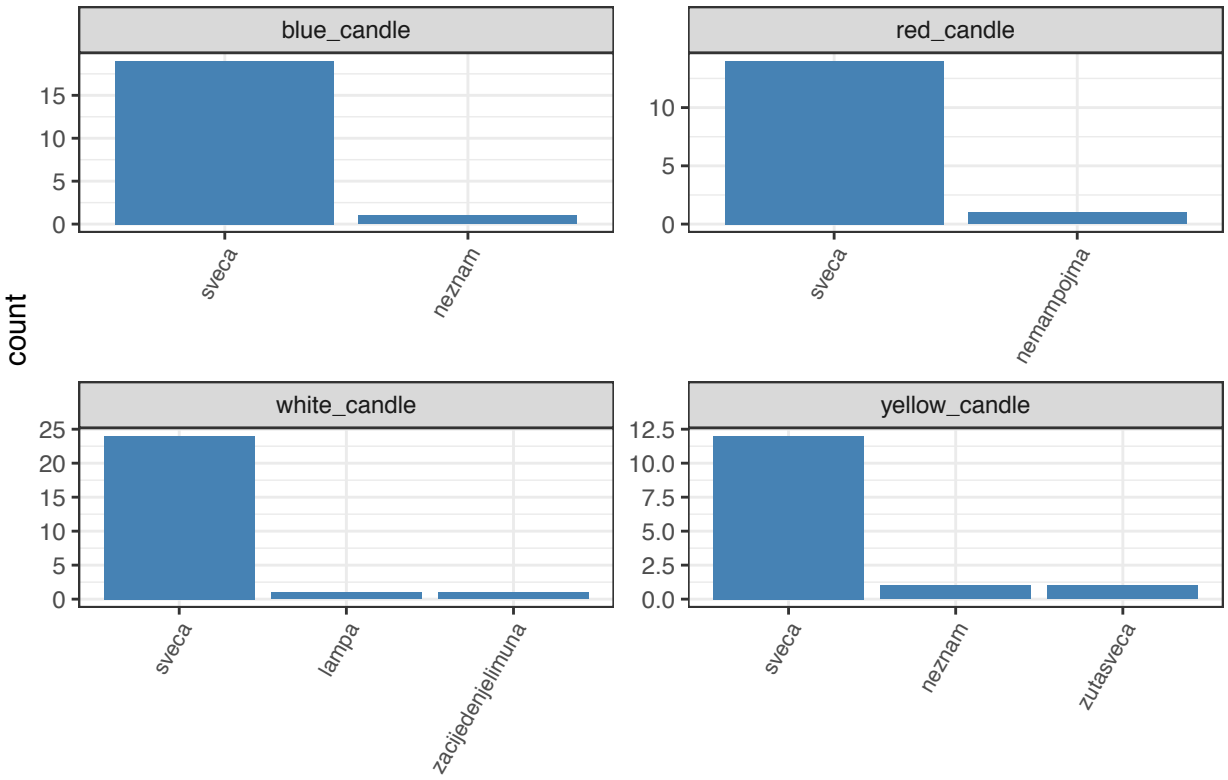
```
##
## [[23]]
```


calendar
black_calendar
green_calendar
orange_calendar
purple_calendar
count
0
5
10
15
20
kalendar
kalendar-mjesecveljaca

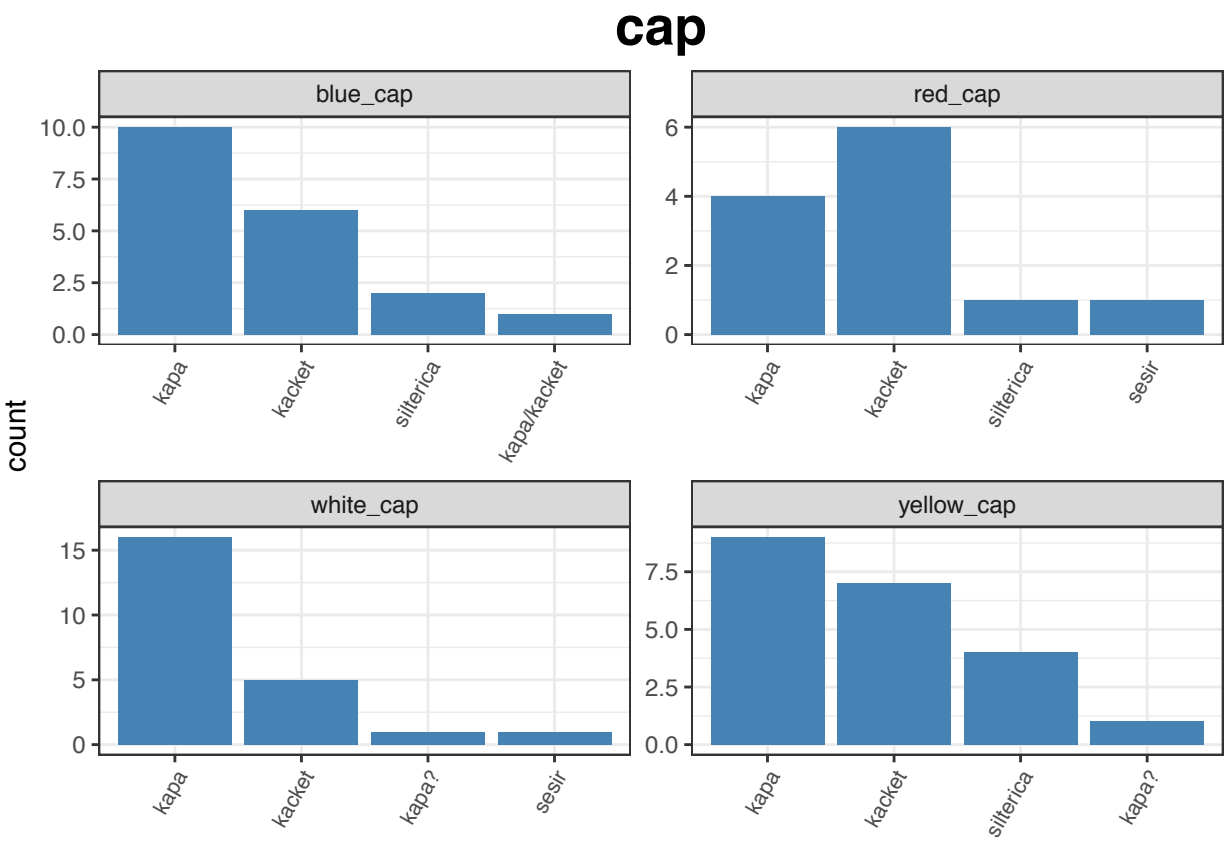
```
##
## [[24]]
```


camera
black_camera
green_camera
orange_camera
purple_camera
count
fotoaparat
kamera
kamera,fotoaparat
kamera?
aparatzaslikanje
polaroid
telefon
fotic
instamatik

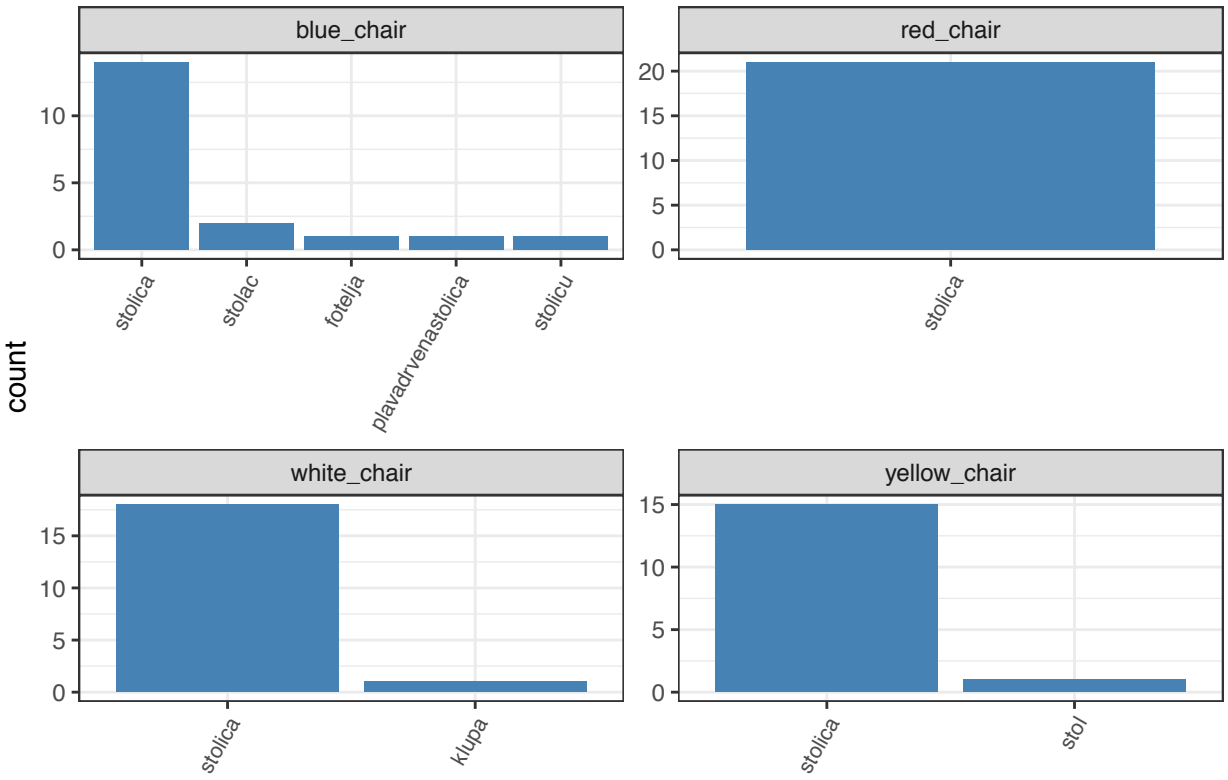
```
##
## [[25]]
```


candle
blue_candle
red_candle
white_candle
yellow_candle
count
sveca
neznam
nemampojma
lampa
zacijedenjelimuna
zutasveca

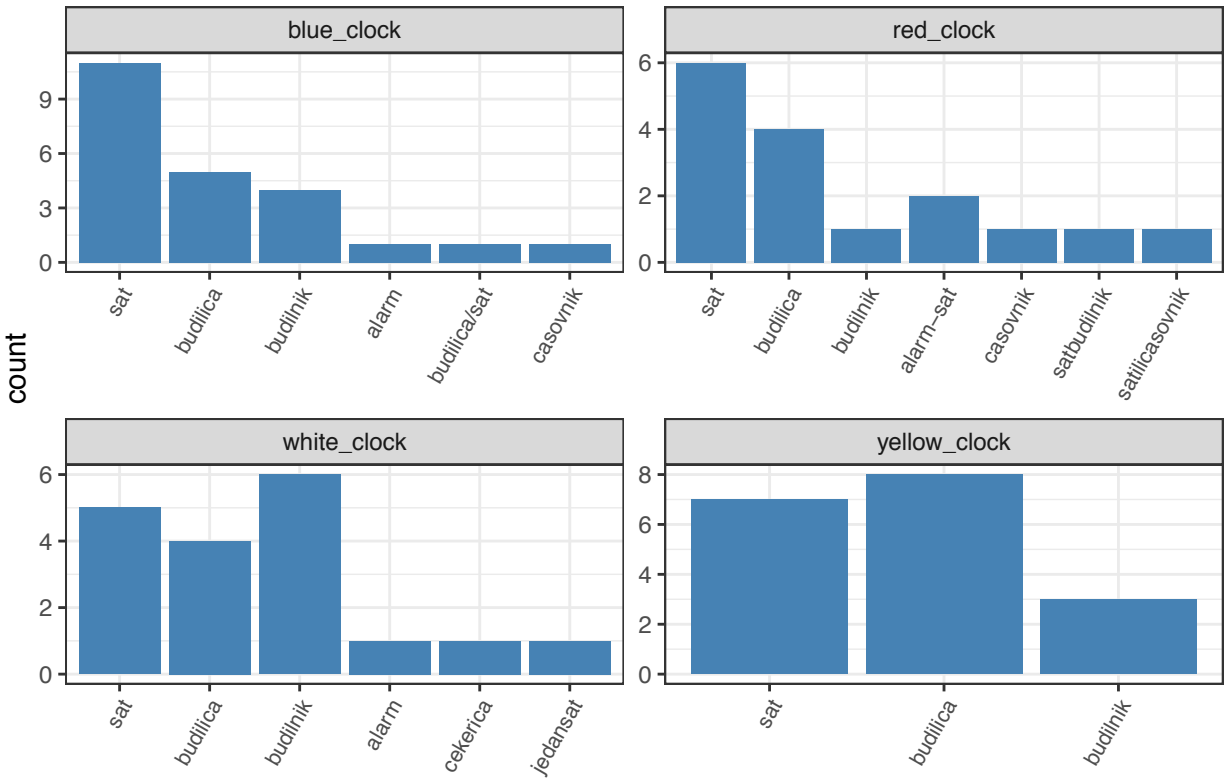
```
##
## [[26]]
```


cap
blue_cap
red_cap
white_cap
yellow_cap
count
kapa
kacket
silterica
kapa/kacket
sesir
kapa?

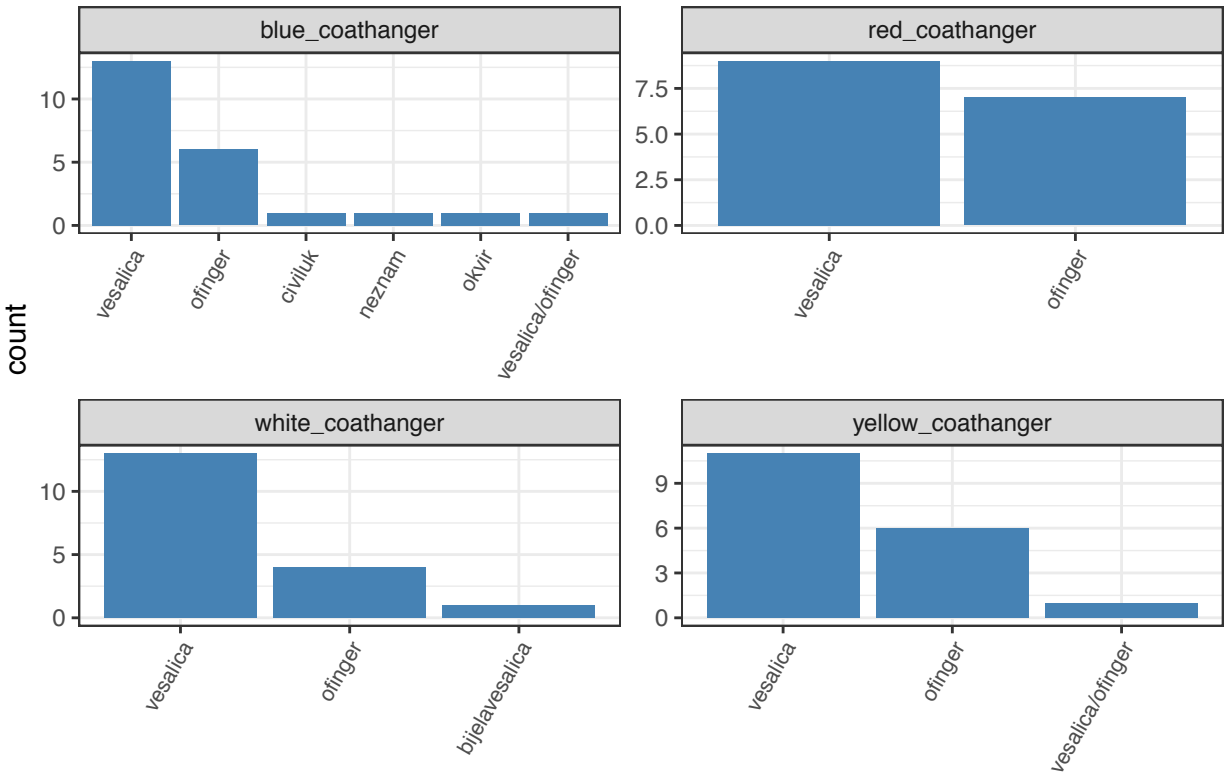
```
##
## [[27]]
```


chair
blue_chair
red_chair
white_chair
yellow_chair
count
0
5
10
15
20
stolica
stolac
fotelja
plavadrvenastolica
stolicu
klupa
stol

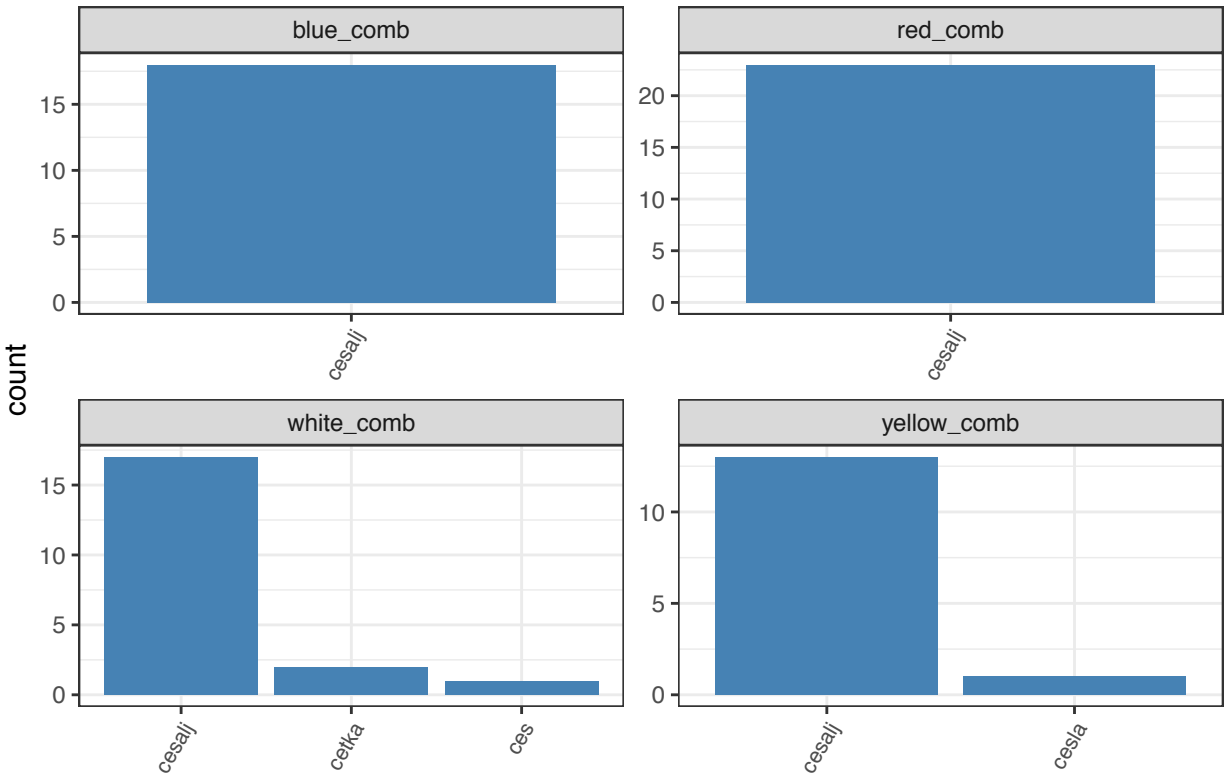
```
##
## [[28]]
```


clock
blue_clock
red_clock
white_clock
yellow_clock
count
sat
budilica
budilnik
alarm
budilica/sat
casovnik
alarm-sat
satbudilnik
satilicasovnik
cekerica
jedansat

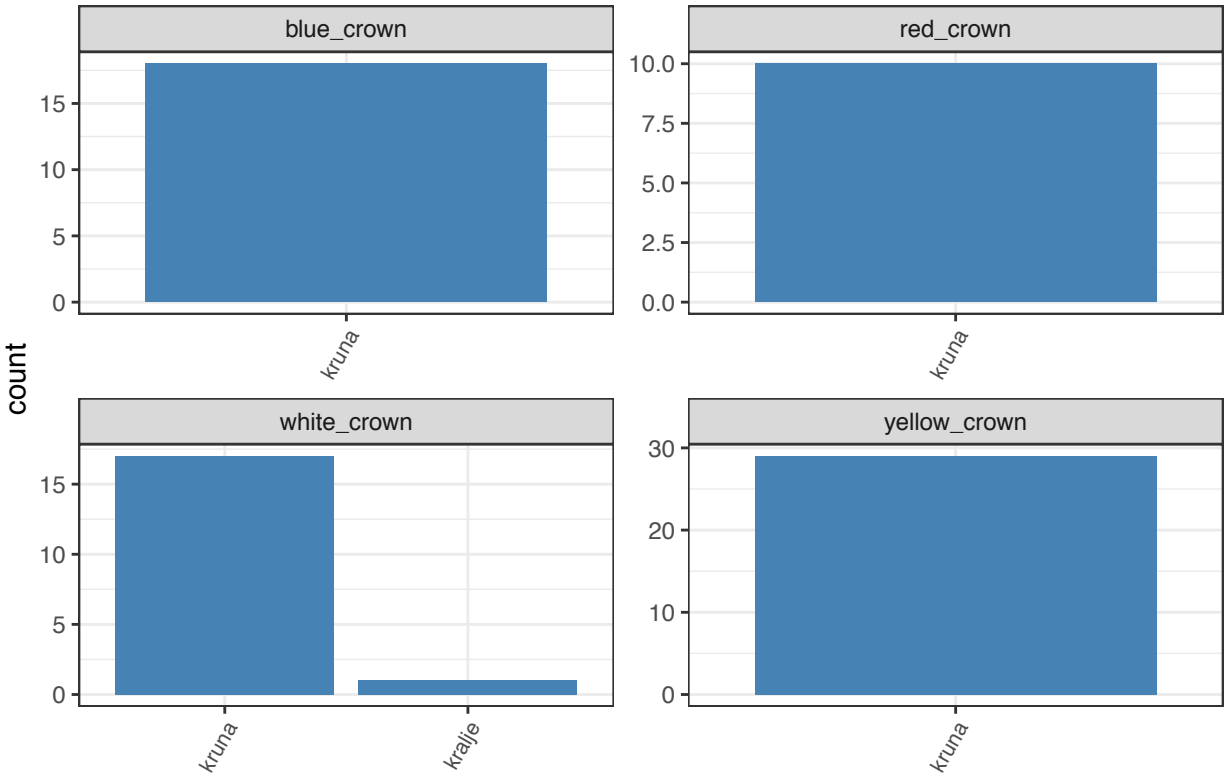
```
##
## [[29]]
```


coathanger
blue_coathanger
red_coathanger
white_coathanger
yellow_coathanger
count
vesalica
ofinger
civiluk
neznam
okvir
vesalica/ofinger
bijelavesalica

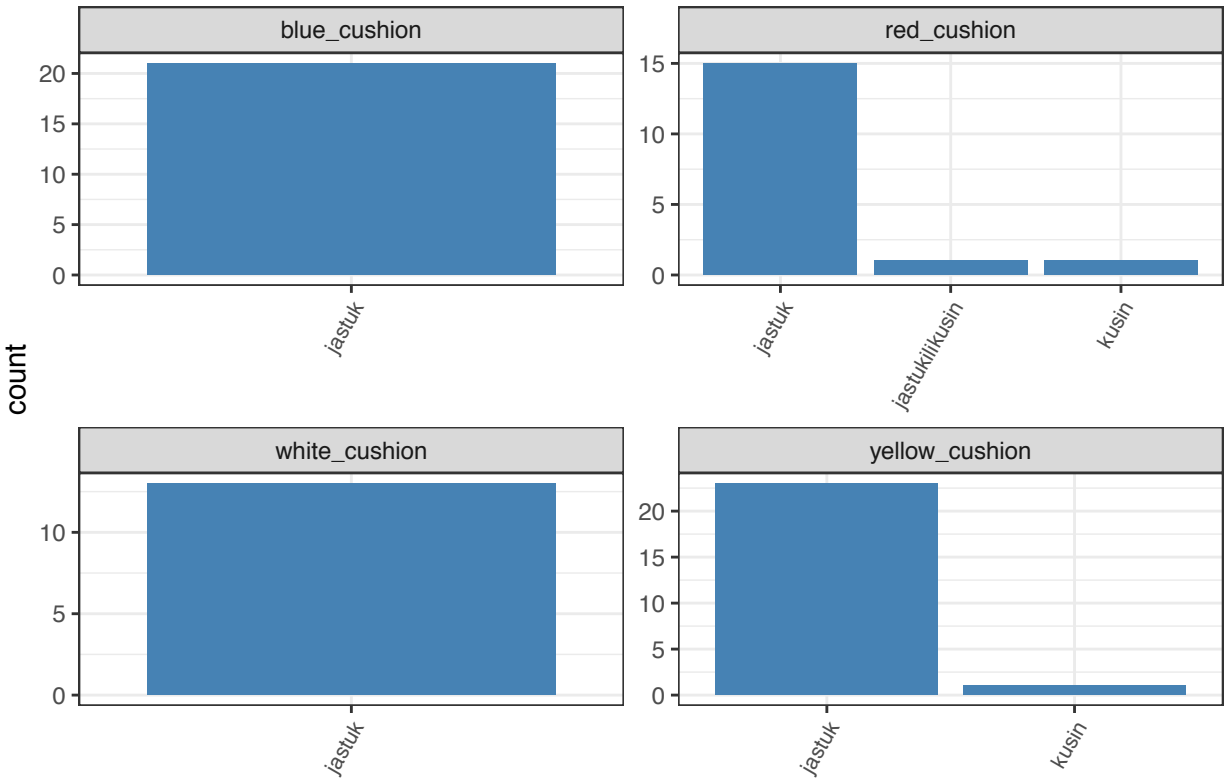
```
##
## [[30]]
```


comb
blue_comb
red_comb
white_comb
yellow_comb
count
cesalj
cetka
ces
cesla

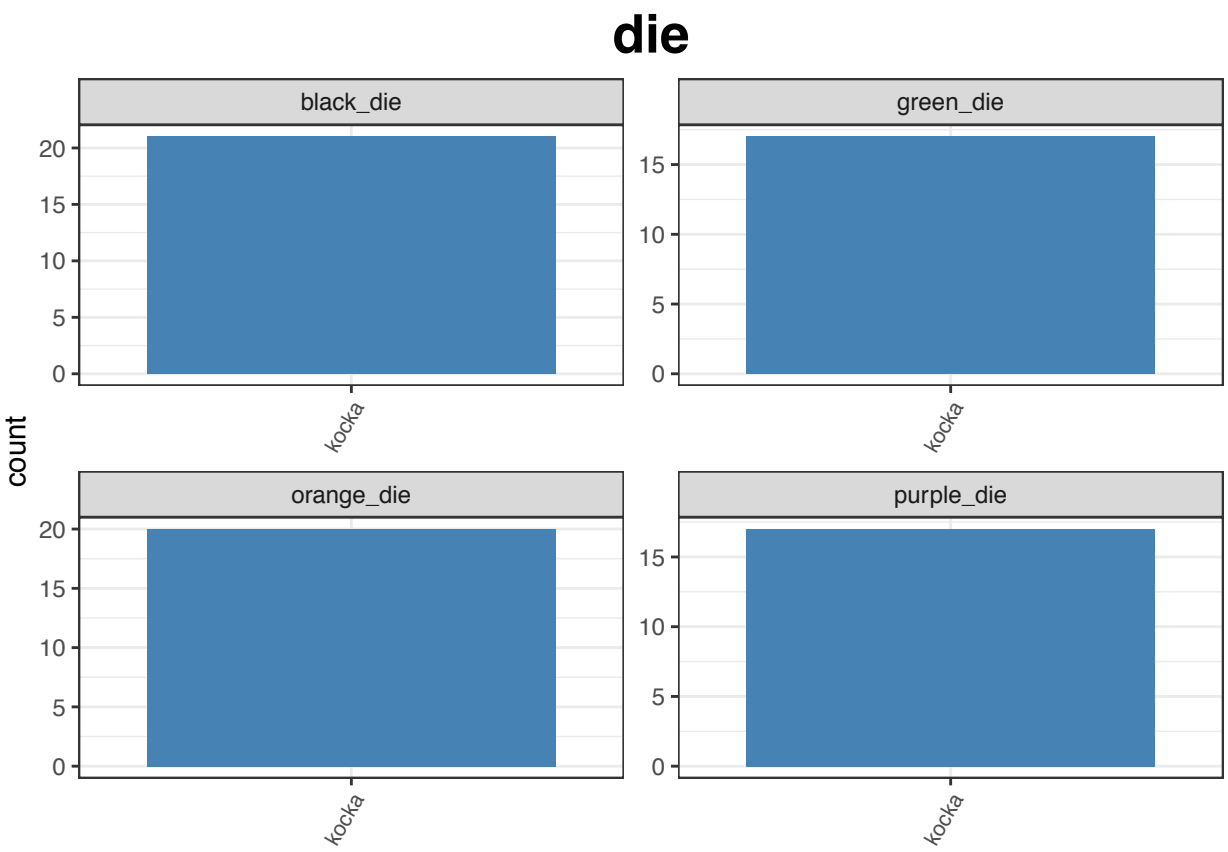
```
##
## [[31]]
```


crown
blue_crown
red_crown
white_crown
yellow_crown
count
kruna
kralje

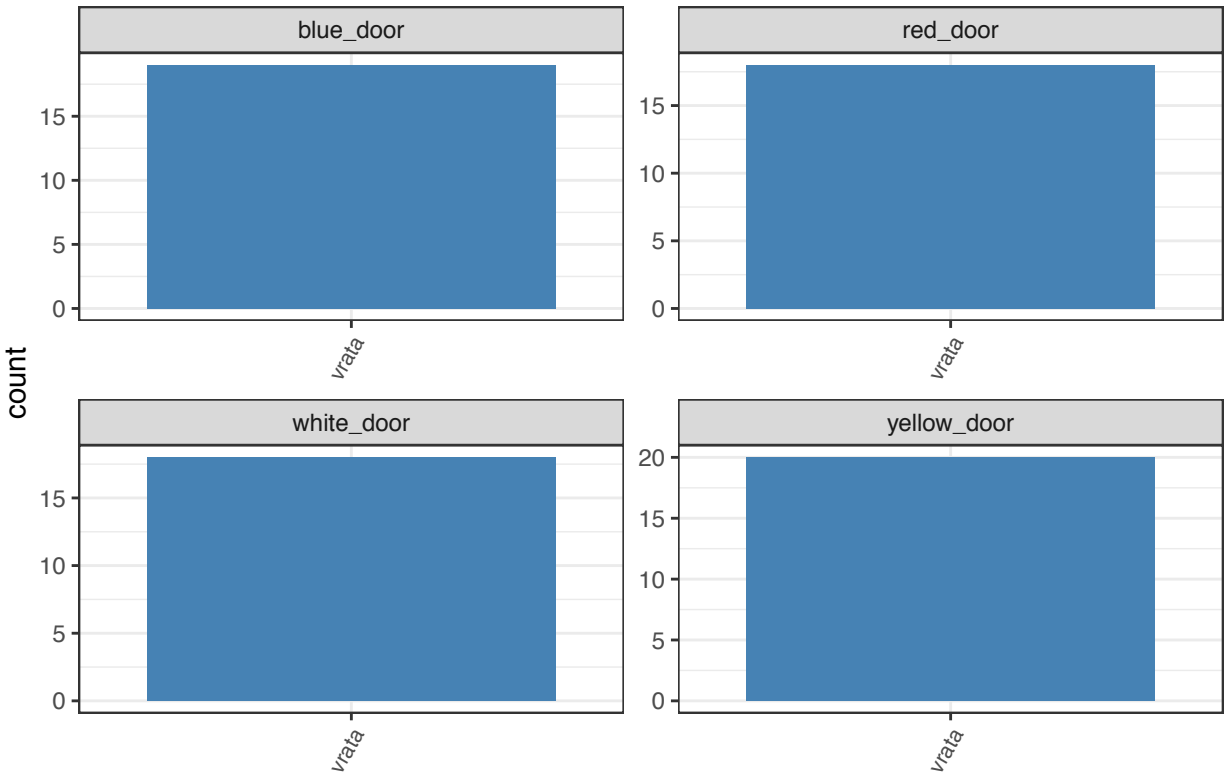
```
##
## [[32]]
```


cushion
blue_cushion
red_cushion
white_cushion
yellow_cushion
count
jastuk
jastukilikusin
kusin

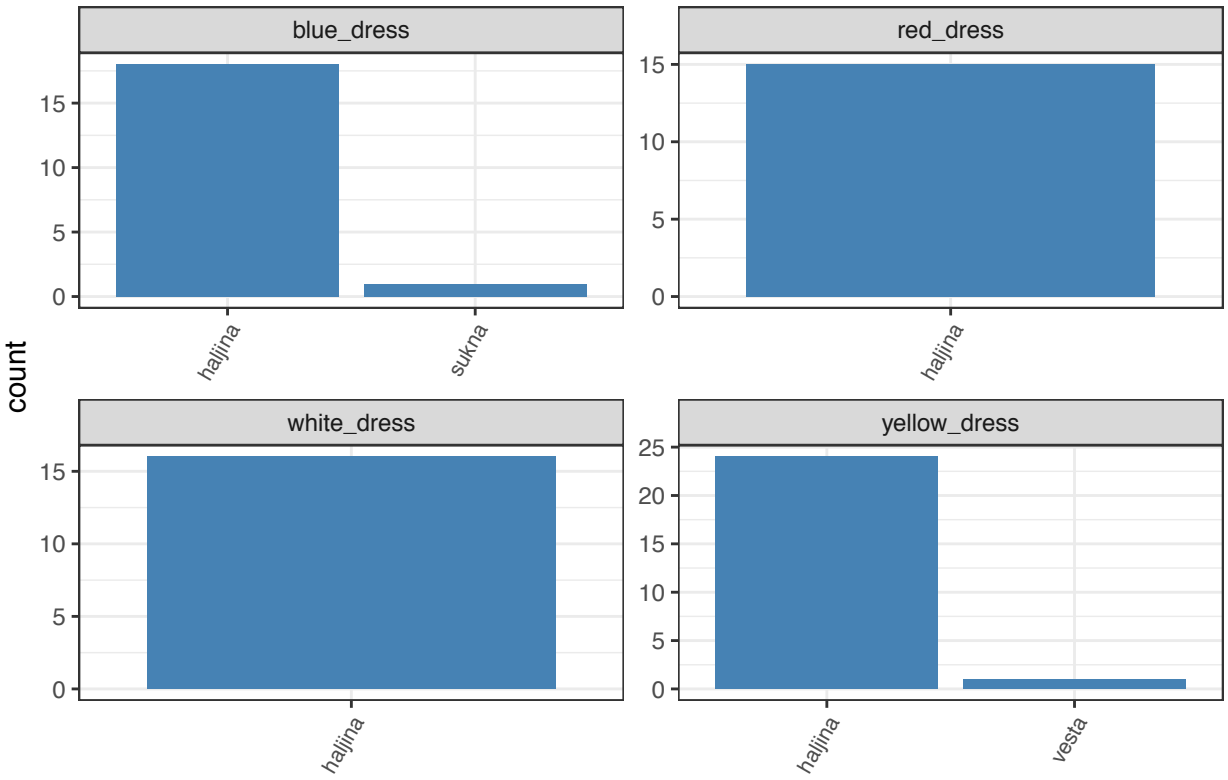
```
##
## [[33]]
```


die
black_die
green_die
orange_die
purple_die
count
kocka
0
5
10
15
20

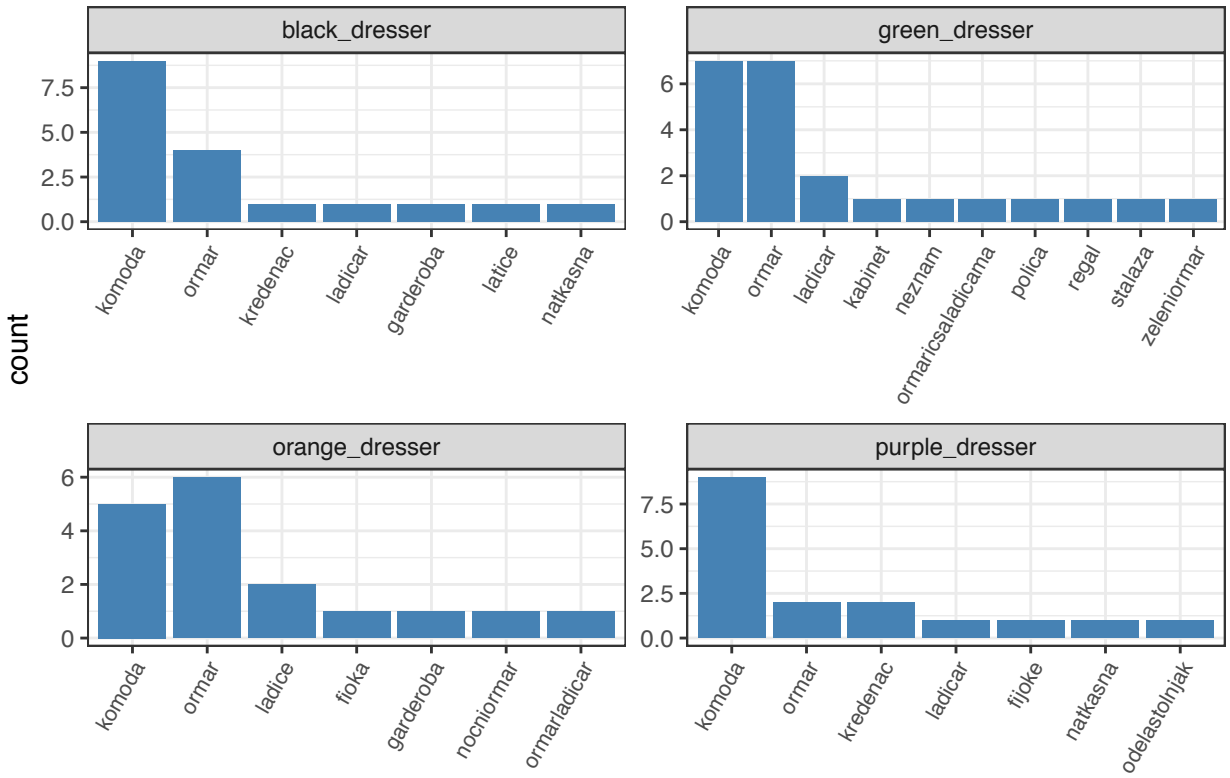
```
##
## [[34]]
```


door
blue_door
red_door
white_door
yellow_door
count
vrata
0
5
10
15
20

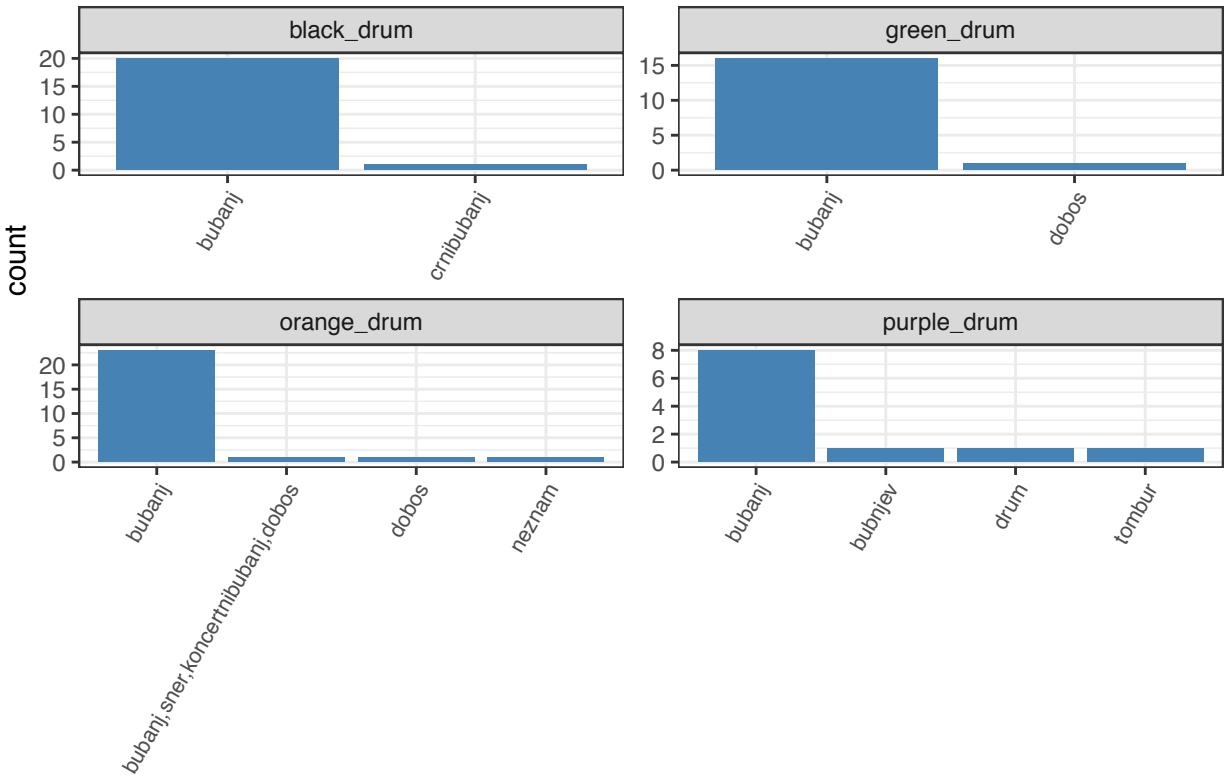
```
##
## [[35]]
```


dress
blue_dress
red_dress
white_dress
yellow_dress
count
haljina
sukna
vesta

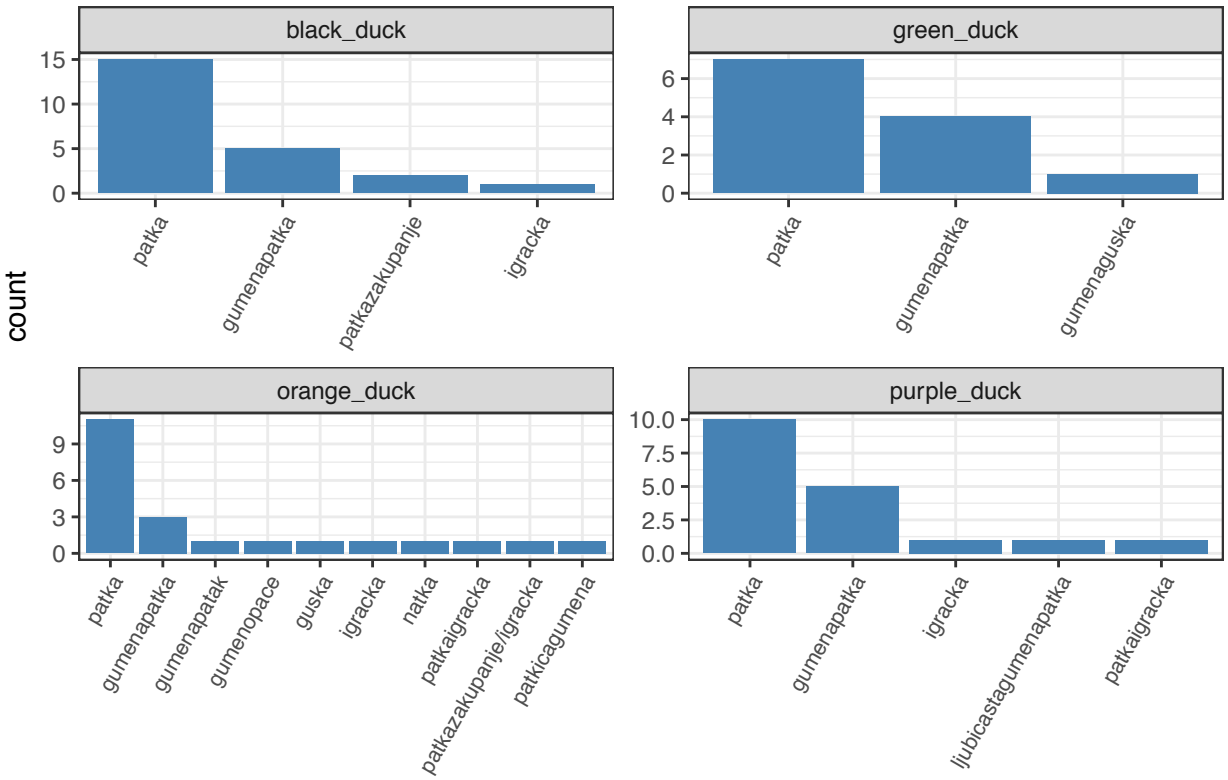
```
##
## [[36]]
```


dresser
black_dresser
green_dresser
orange_dresser
purple_dresser
count
komoda
ormar
kredenac
ladicar
garderoba
latice
natkasna
ladicar
kabinet
neznam
ormaricsaladicama
polica
regal
stalaza
zeleniormar
ladice
fioka
nocniormar
ormarladicar
fijoke
odelastolnjak

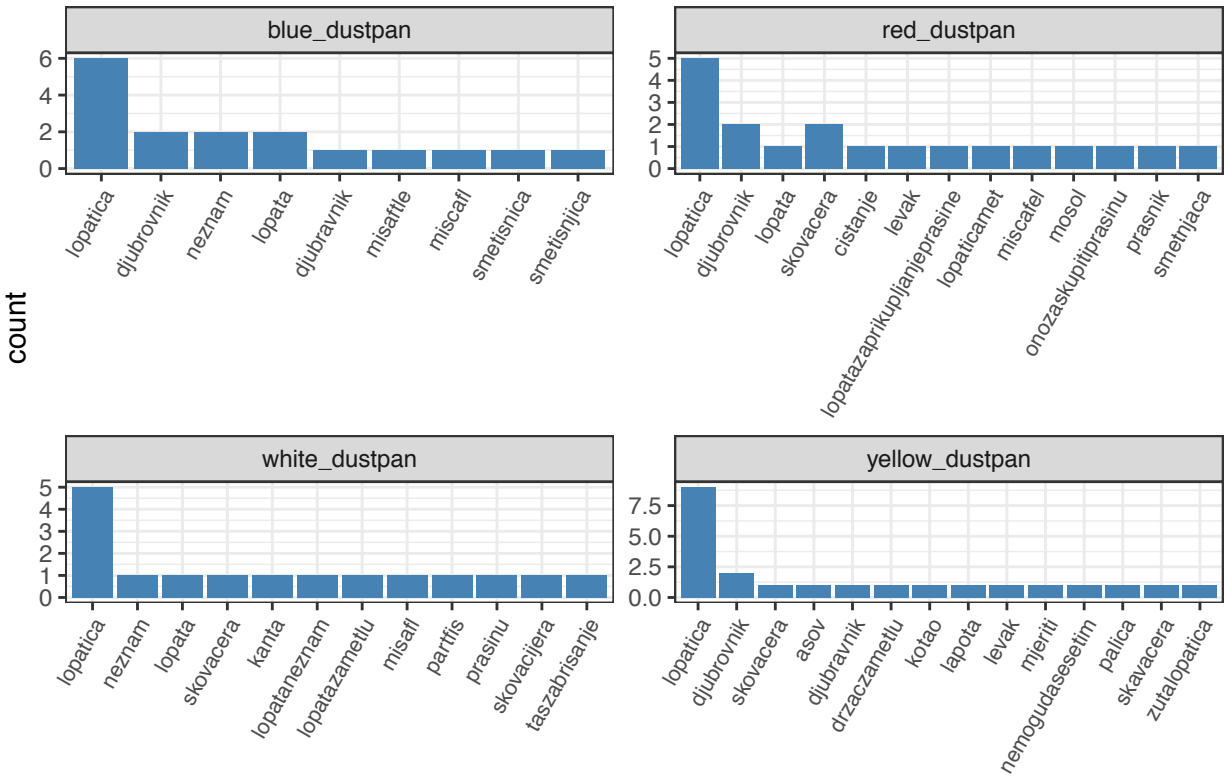
```
##
## [[37]]
```


drum
black_drum
green_drum
orange_drum
purple_drum
count
bubanj
crnibubanj
dobos
bubanj;sner;koncertnibubanj;dobos
neznam
bubnjev
drum
tombur

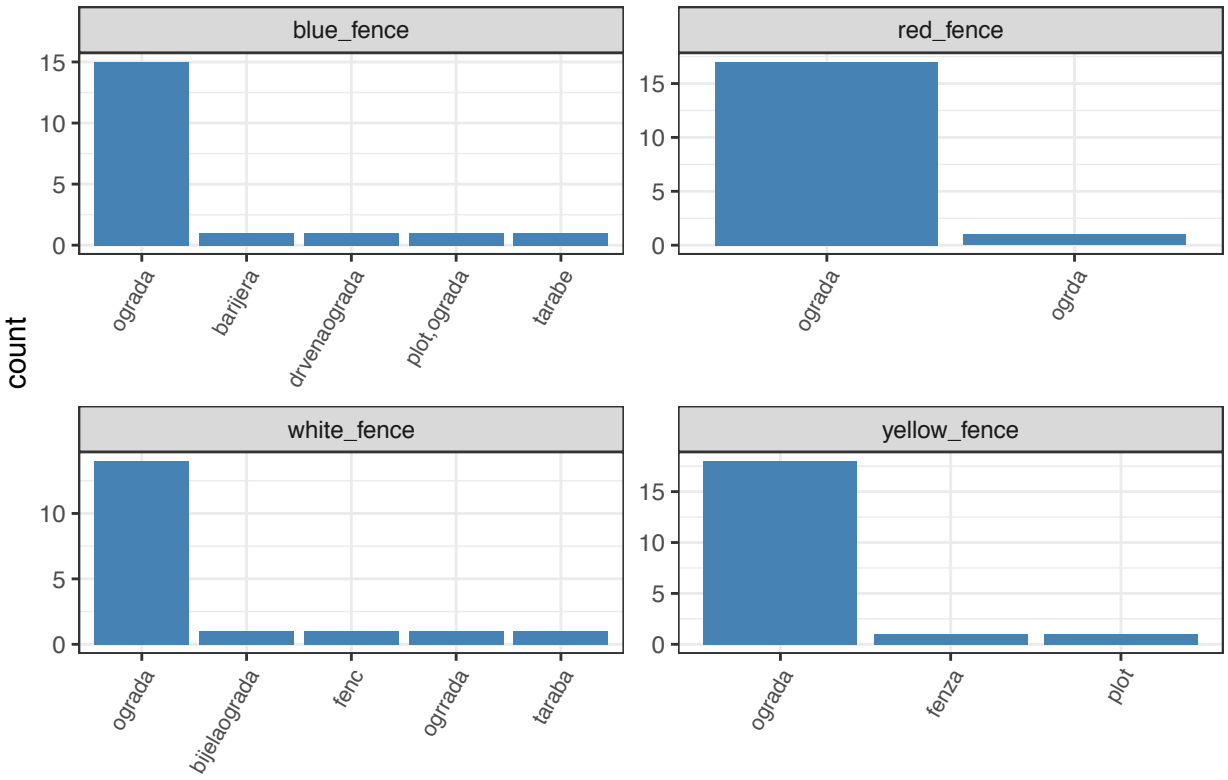
```
##
## [[38]]
```


duck
black_duck
green_duck
orange_duck
purple_duck
count
patka
gumenapatka
patkazakupanje
igracka
gumenaguska
guska
natka
patkaigracka
patkazakupanje/igracka
patkicagumena
gumenapatak
gumenopace
ljubicastagumenapatka

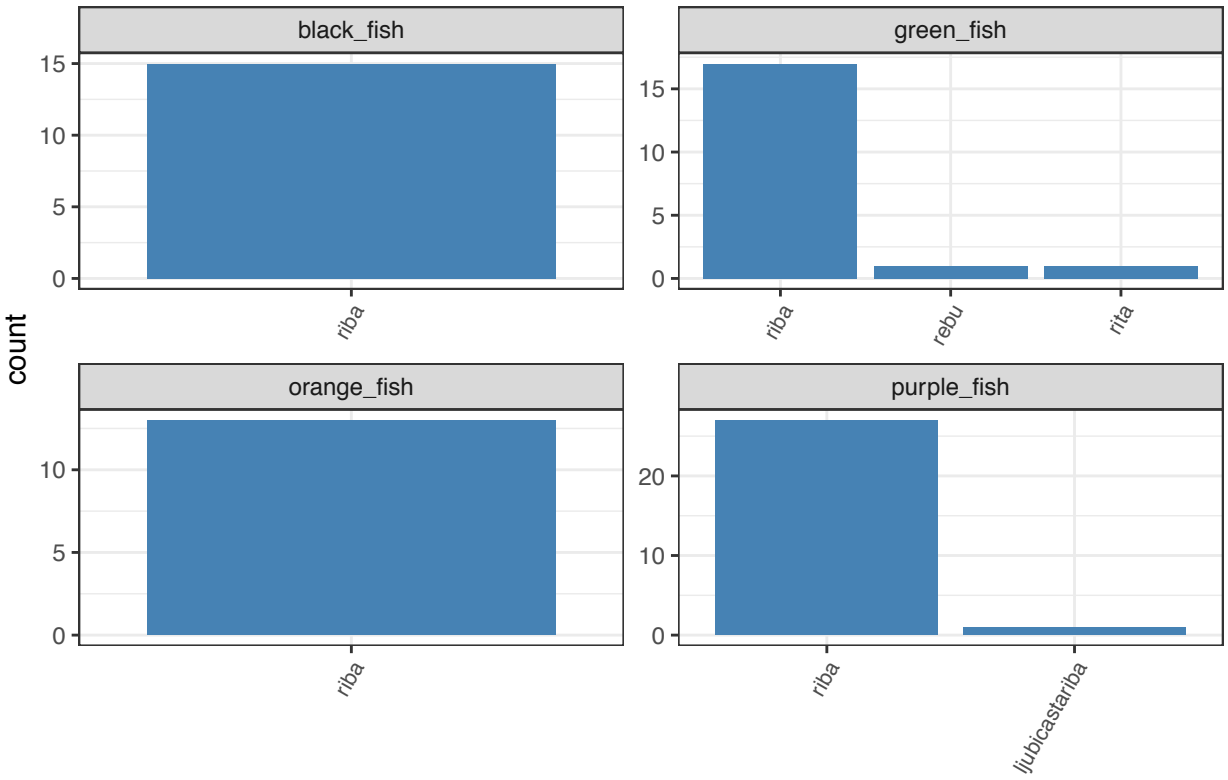
```
##
## [[39]]
```


dustpan
blue_dustpan
red_dustpan
white_dustpan
yellow_dustpan
count
lopatica
djubrovnik
neznam
lopata
djubravnik
misafle
miscafl
smetisnica
smetisnjica
lopatazaprikupljanjeprasine
lopaticamet
miscafel
mosol
onozaskupitiprasinu
prasnik
smetnjaca
kanta
lopataneznam
lopatazametlu
misafl
partfis
prasinu
skovacijera
taszabrisanje
asov
drzaczametlu
kotao
lapota
levak
mjeriti
nemogudasesetim
palica
skavacera
zutalopatica

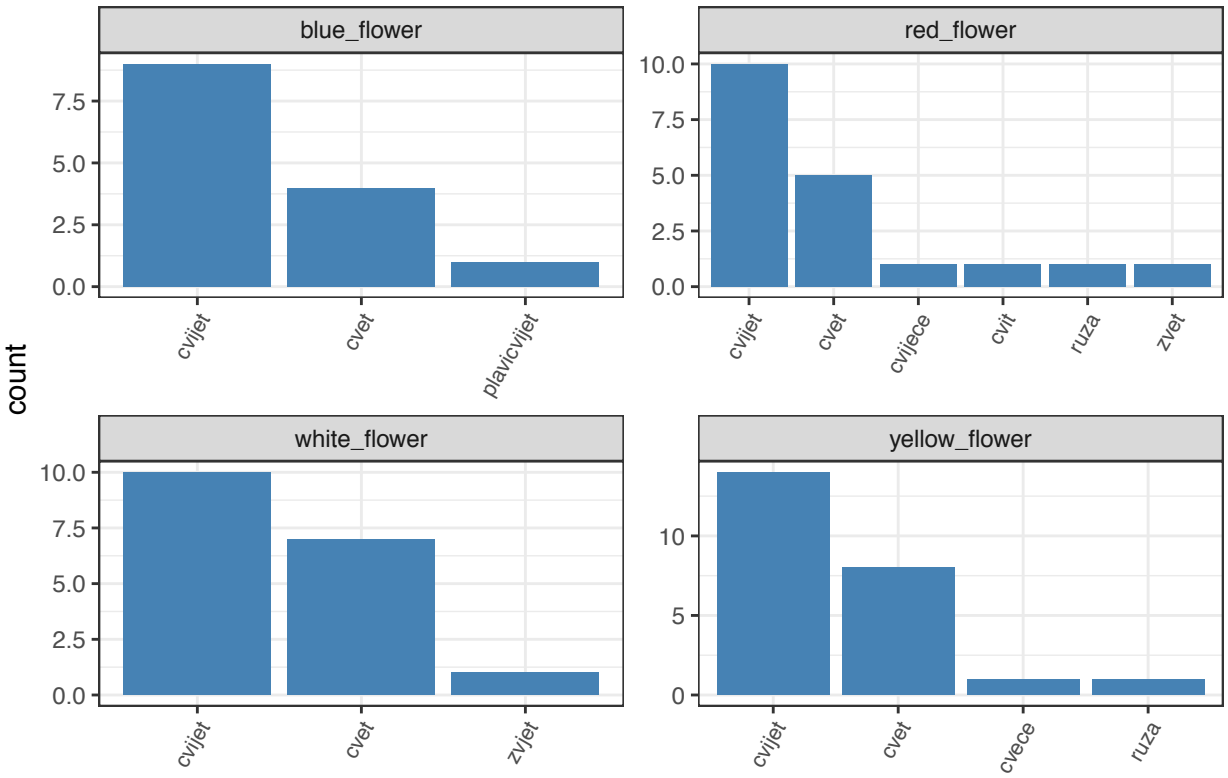
```
##
## [[40]]
```


fence
blue_fence
red_fence
white_fence
yellow_fence
count
0
5
10
15
ograda
barijera
drvenaograda
plot,ograda
tarabe
ograda
ogrda
ograda
bijelaograda
fenc
ogrrada
taraba
ograda
fenza
plot

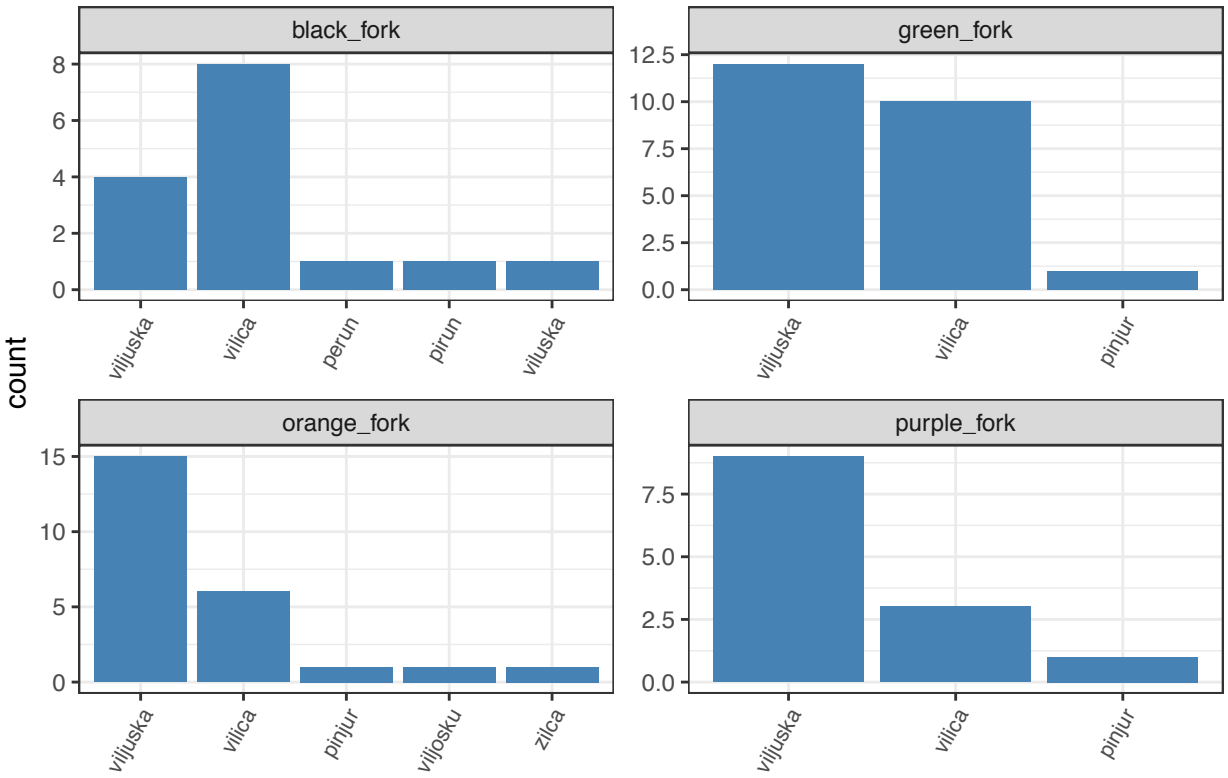
```
##
## [[41]]
```


fish
black_fish
green_fish
orange_fish
purple_fish
count
15
10
5
0
20
riba
rebu
rita
ljubicastariba

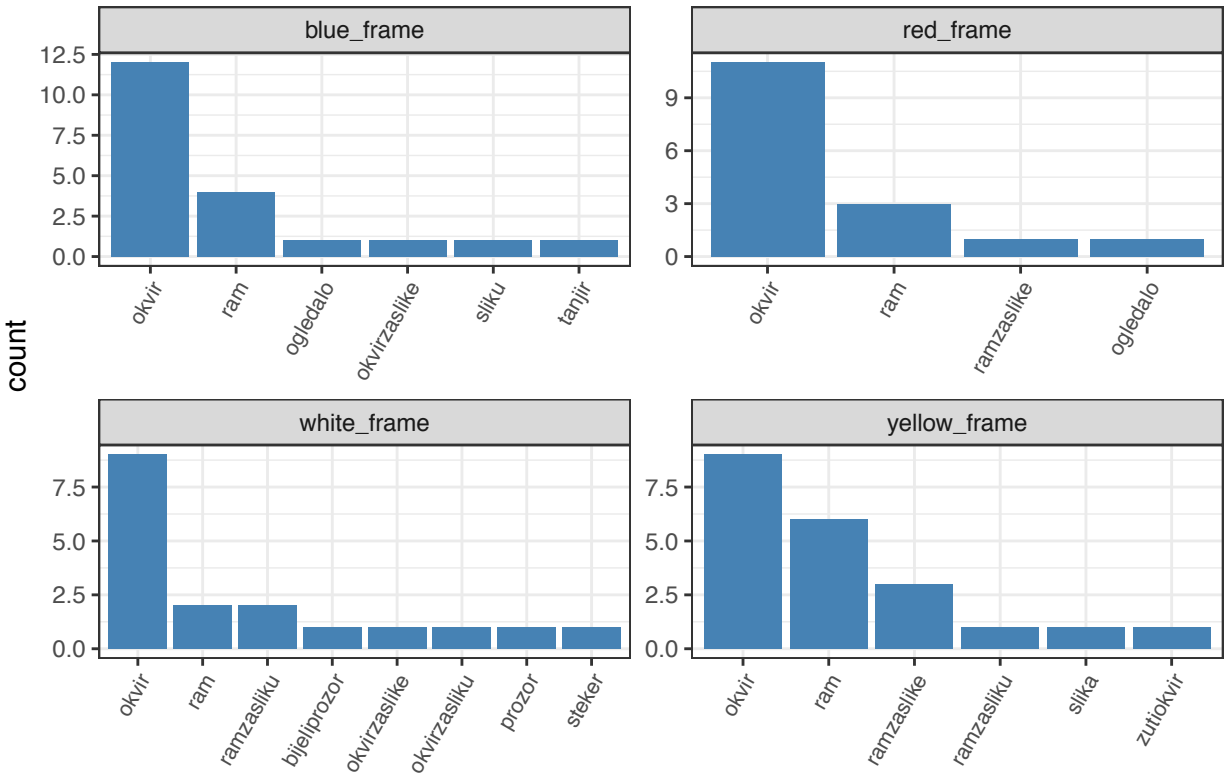
```
##
## [[42]]
```


flower
blue_flower
red_flower
white_flower
yellow_flower
count
cvijet
cvet
plavicvijet
cvijece
cvit
ruza
zvet
zvjet
cvece

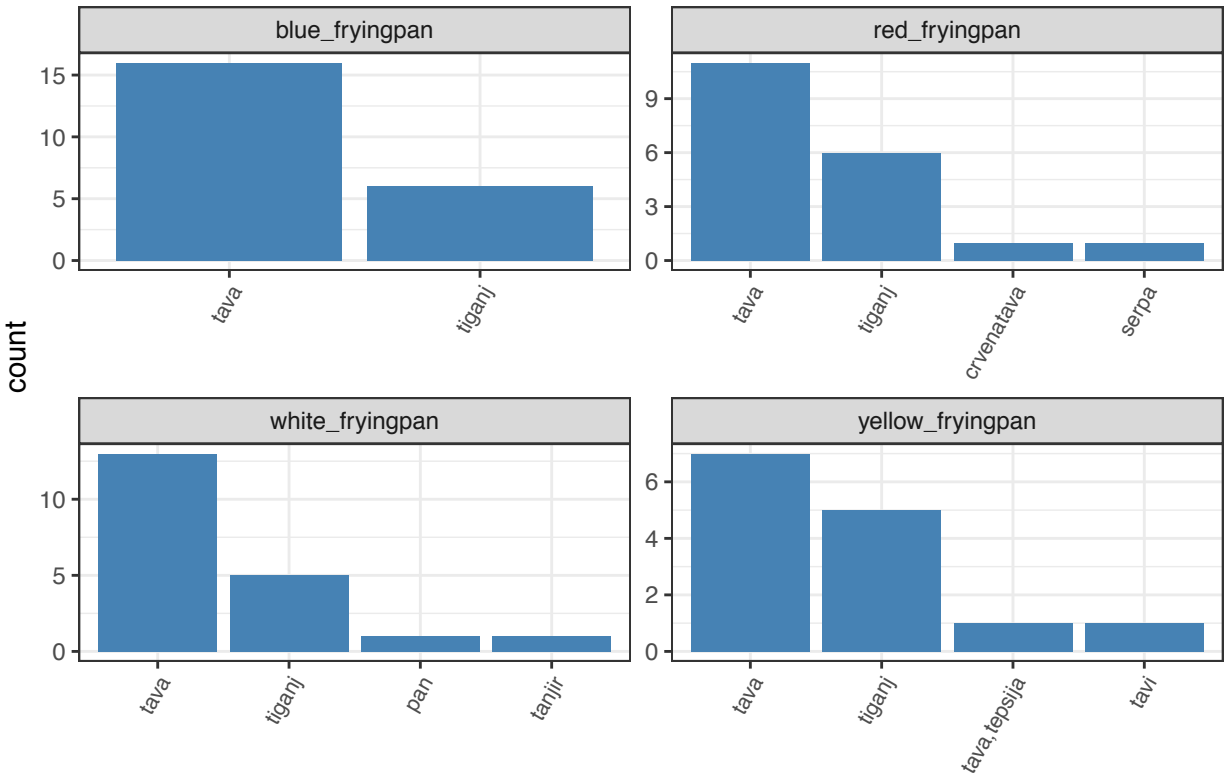
```
##
## [[43]]
```


fork
black_fork
green_fork
orange_fork
purple_fork
count
viljuska
vilica
perun
pirun
viluska
pinjur
viljosku
zilca

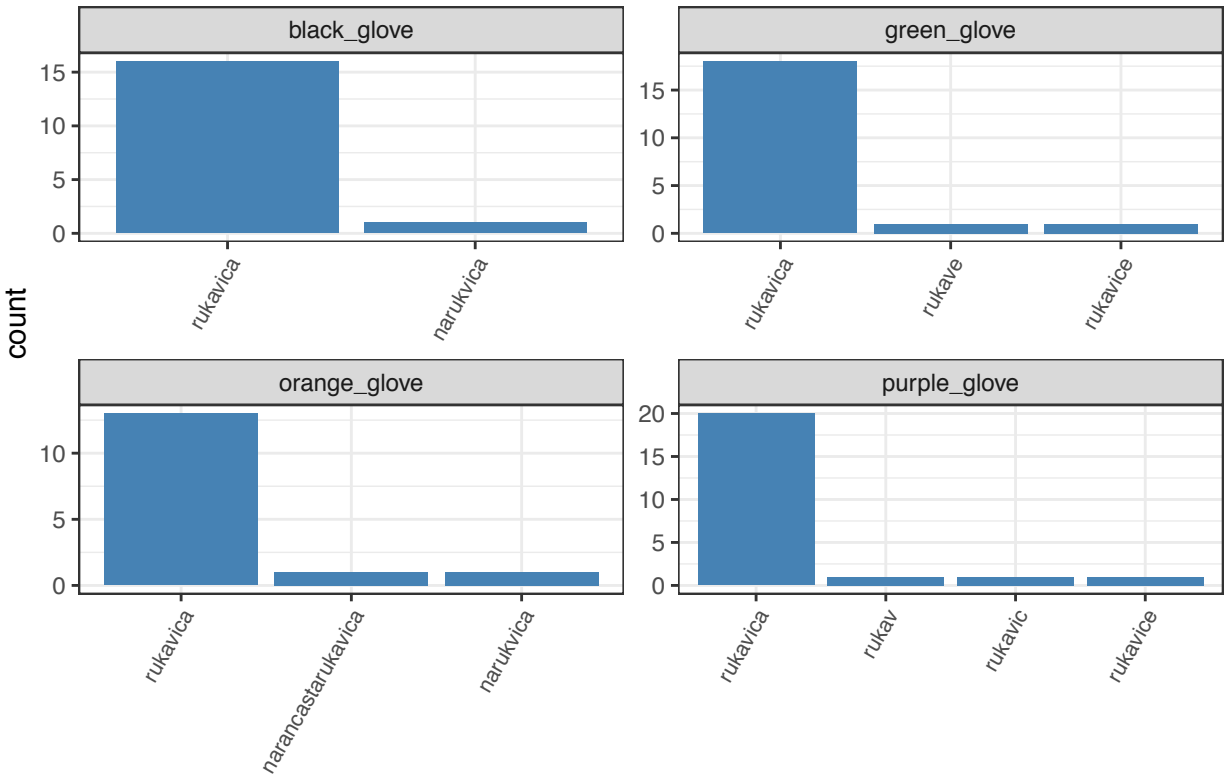
```
##
## [[44]]
```


frame
blue_frame
red_frame
white_frame
yellow_frame
count
okvir
ram
ogledalo
okvirzaslike
sliku
tanjir
ramzaslike
ramzasliku
bijeliprozor
prozor
steker
slika
zutiokvir

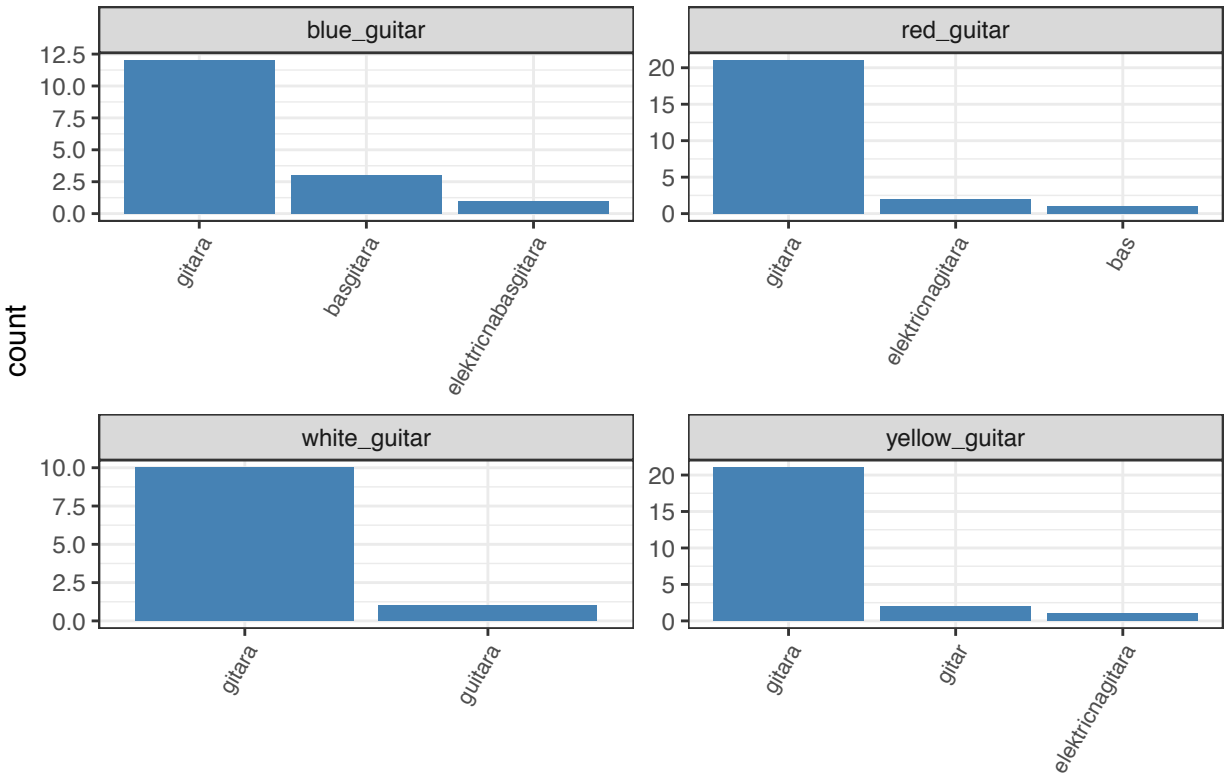
```
##
## [[45]]
```


fryingpan
blue_fryingpan
red_fryingpan
white_fryingpan
yellow_fryingpan
count
tava
tiganj
crvenatava
serpa
pan
tanjir
tava,tepsija
tavi

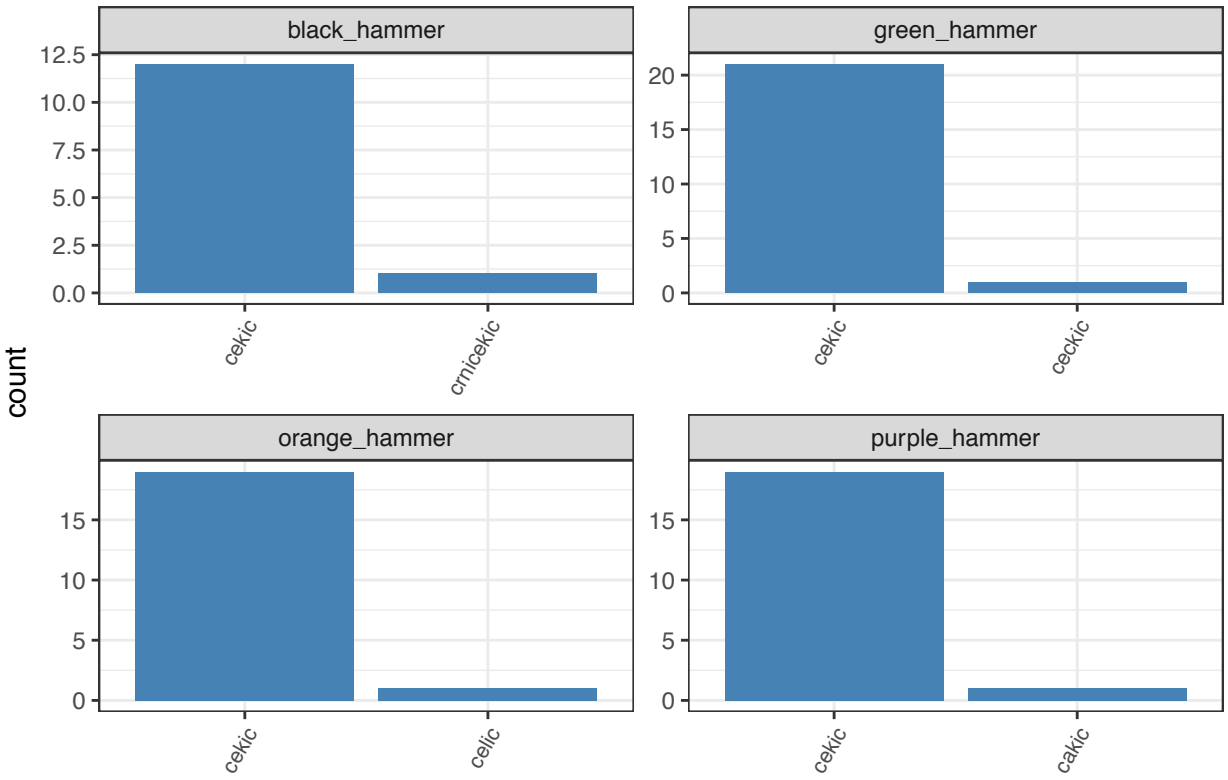
```
##
## [[46]]
```


glove
black_glove
green_glove
orange_glove
purple_glove
count
rukavica
narukvica
rukave
rukavice
narancastarukavica
rukav
rukavic

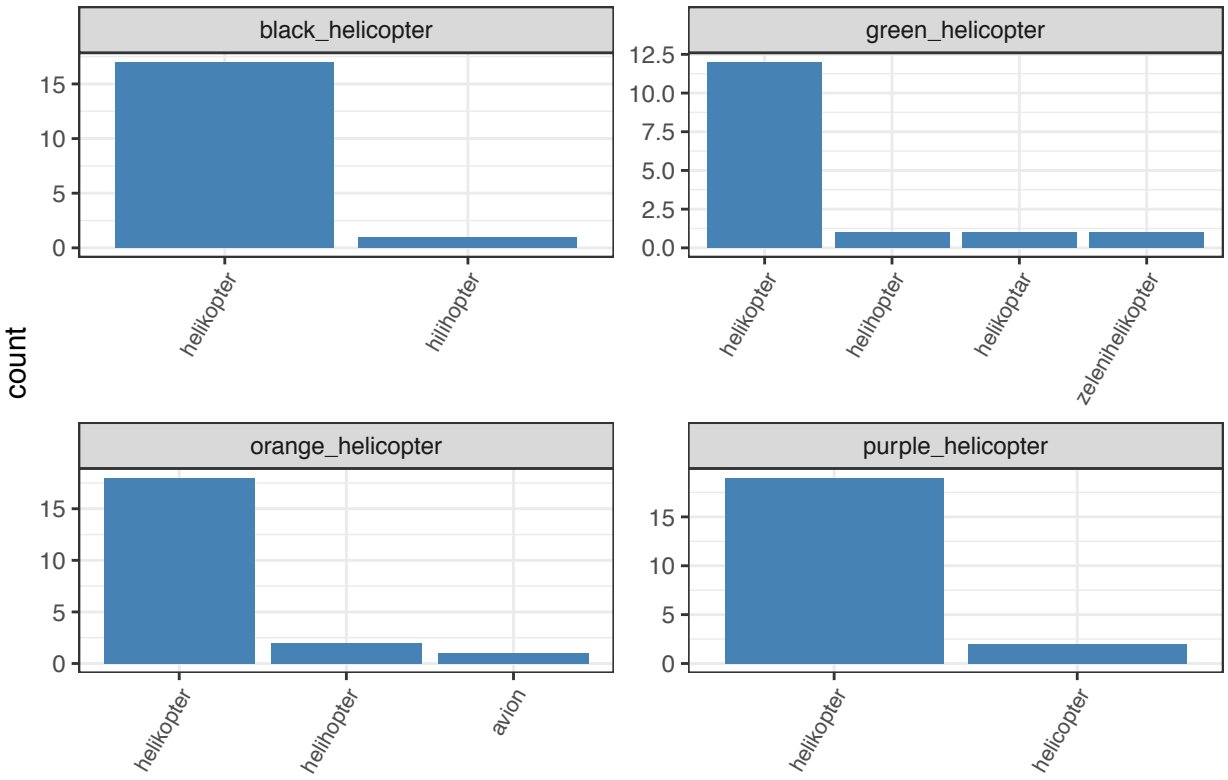
```
##
## [[47]]
```


guitar
blue_guitar
red_guitar
white_guitar
yellow_guitar
count
gitara
basgitara
elektricnabasgitara
elektricnagitara
bas
guitara
gitar

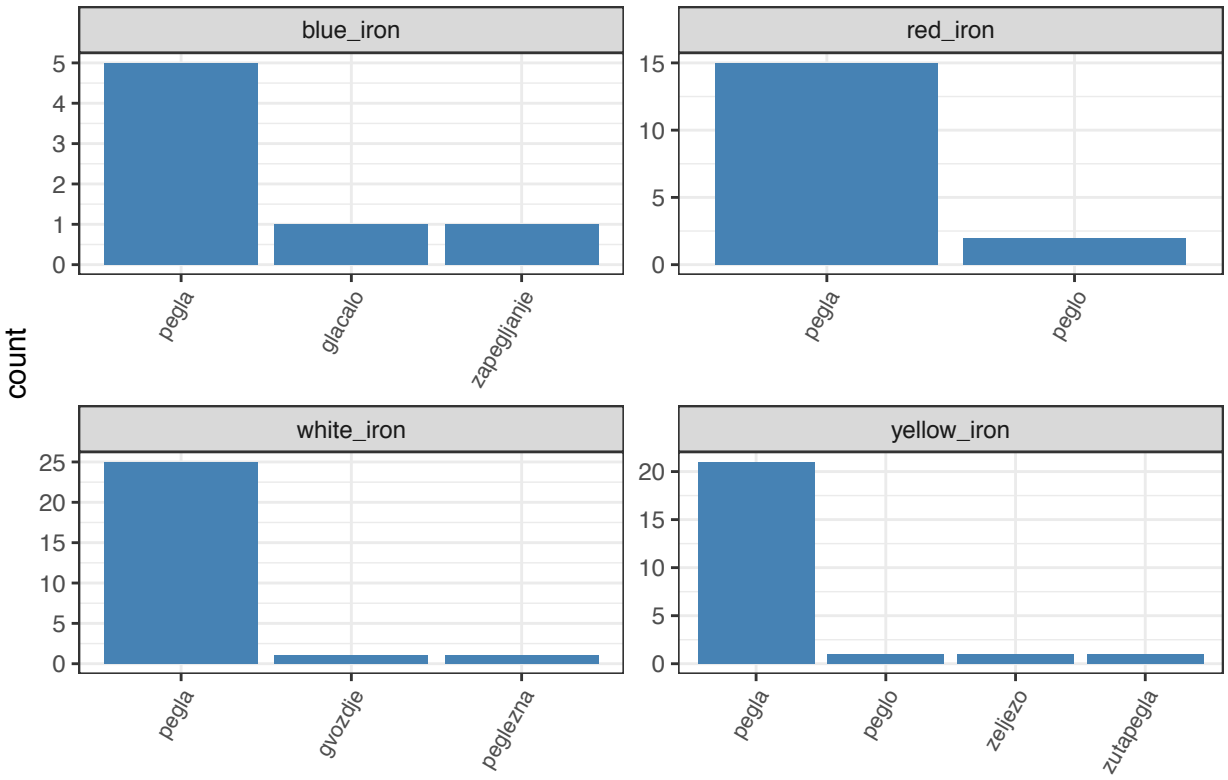
```
##
## [[48]]
```


hammer
black_hammer
green_hammer
orange_hammer
purple_hammer
count
12.5
10.0
7.5
5.0
2.5
0.0
20
15
10
5
0
cekic
crnicekic
ceckic
celic
cakic

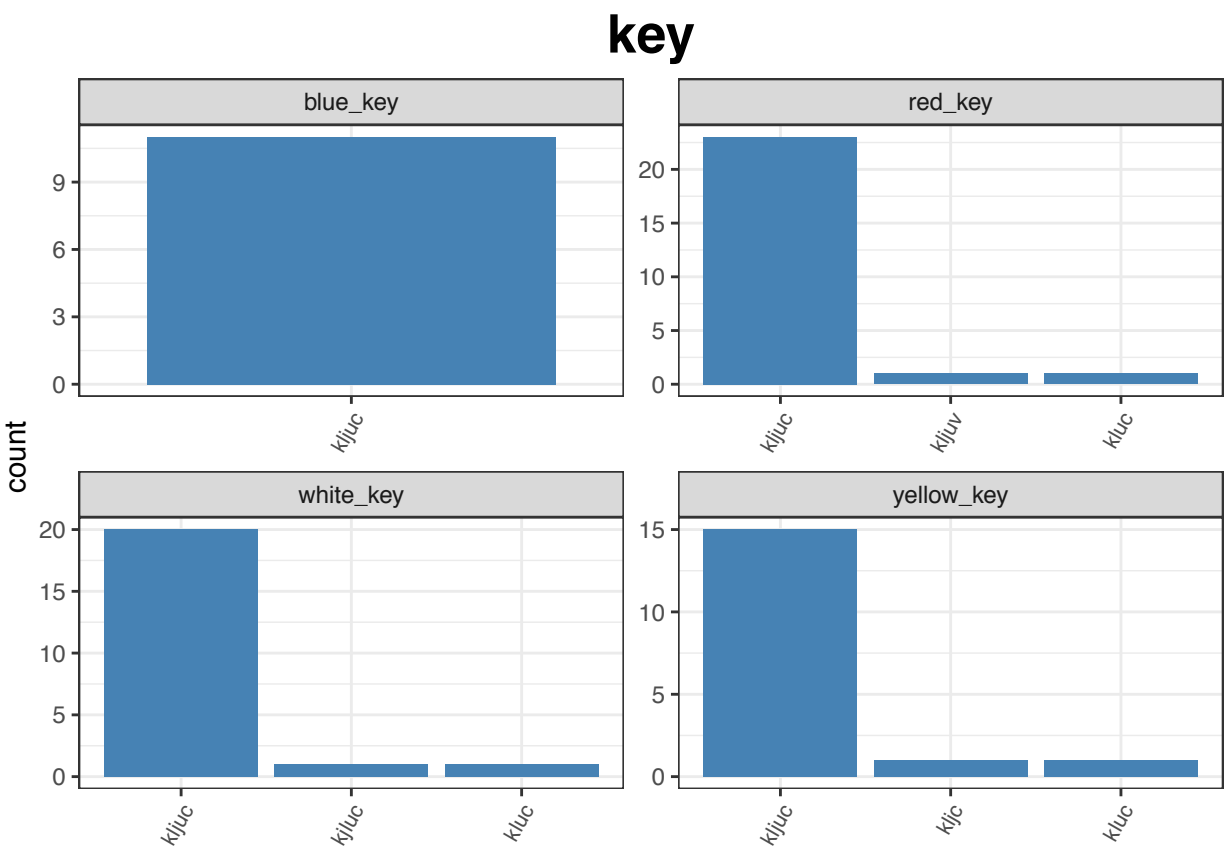
```
##
## [[49]]
```


helicopter
black_helicopter
green_helicopter
orange_helicopter
purple_helicopter
count
helikopter
hiiihopter
helihopter
helikoptar
zelenihelikopter
avion
helicopter

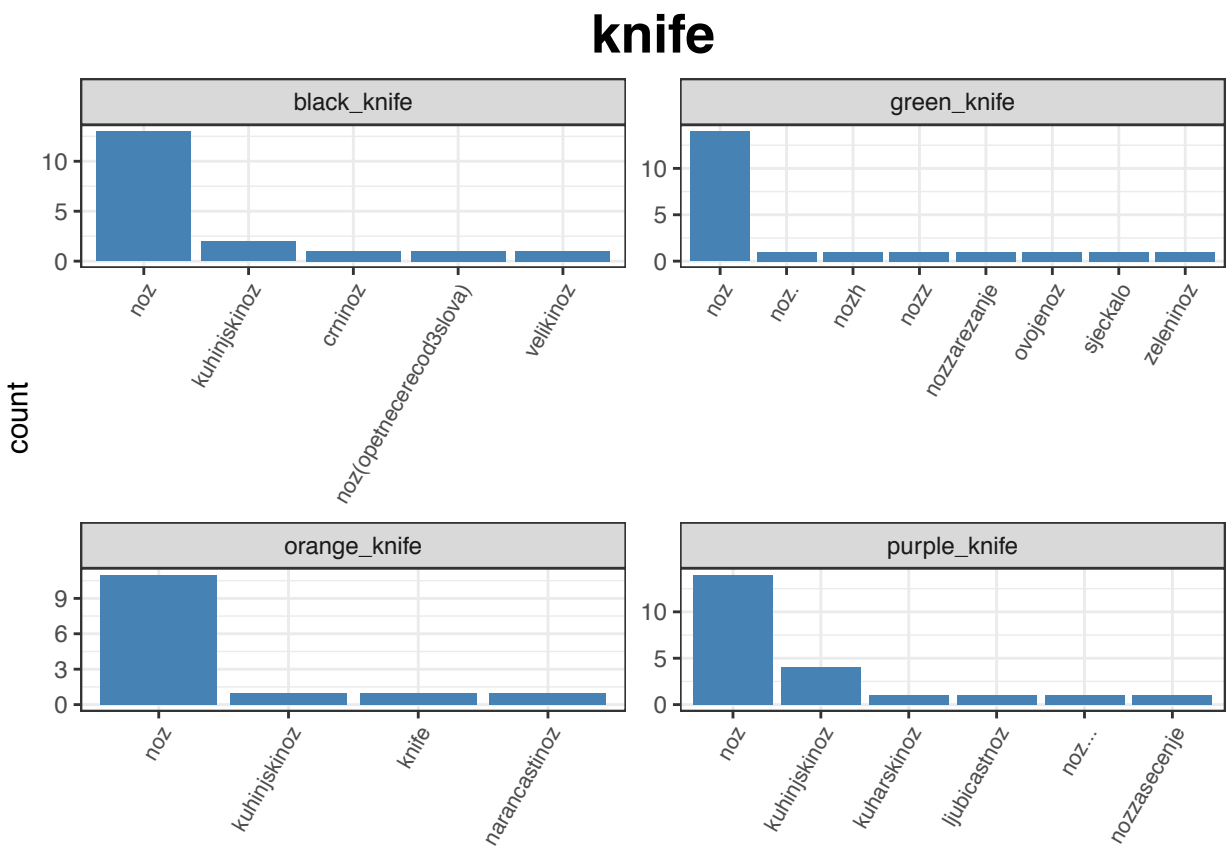
```
##
## [[50]]
```


iron
blue_iron
red_iron
white_iron
yellow_iron
count
pegla
glacalo
zapegljanje
peglo
gvozdje
peglezna
zeljezo
zutapegla

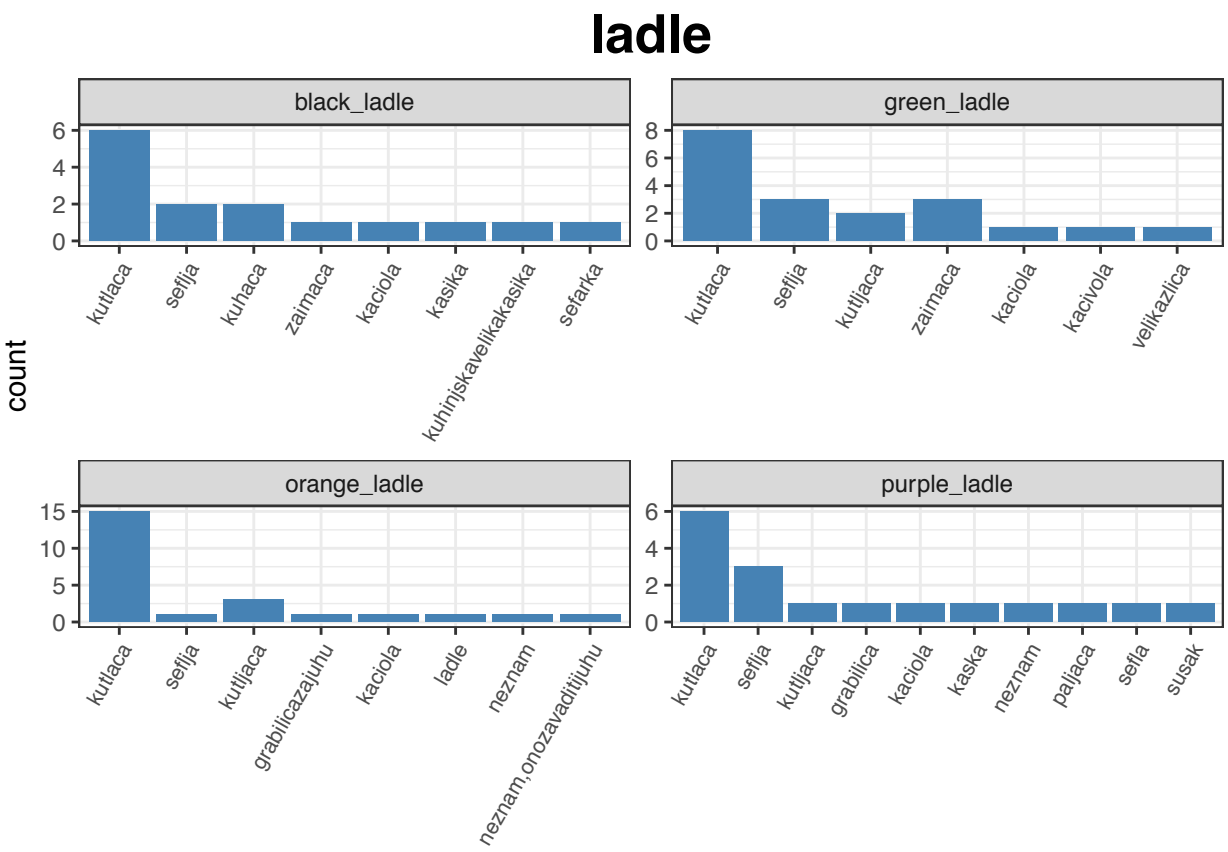
```
##
## [[51]]
```


key
blue_key
red_key
white_key
yellow_key
count
kljuc
kljuv
kluc
kljc

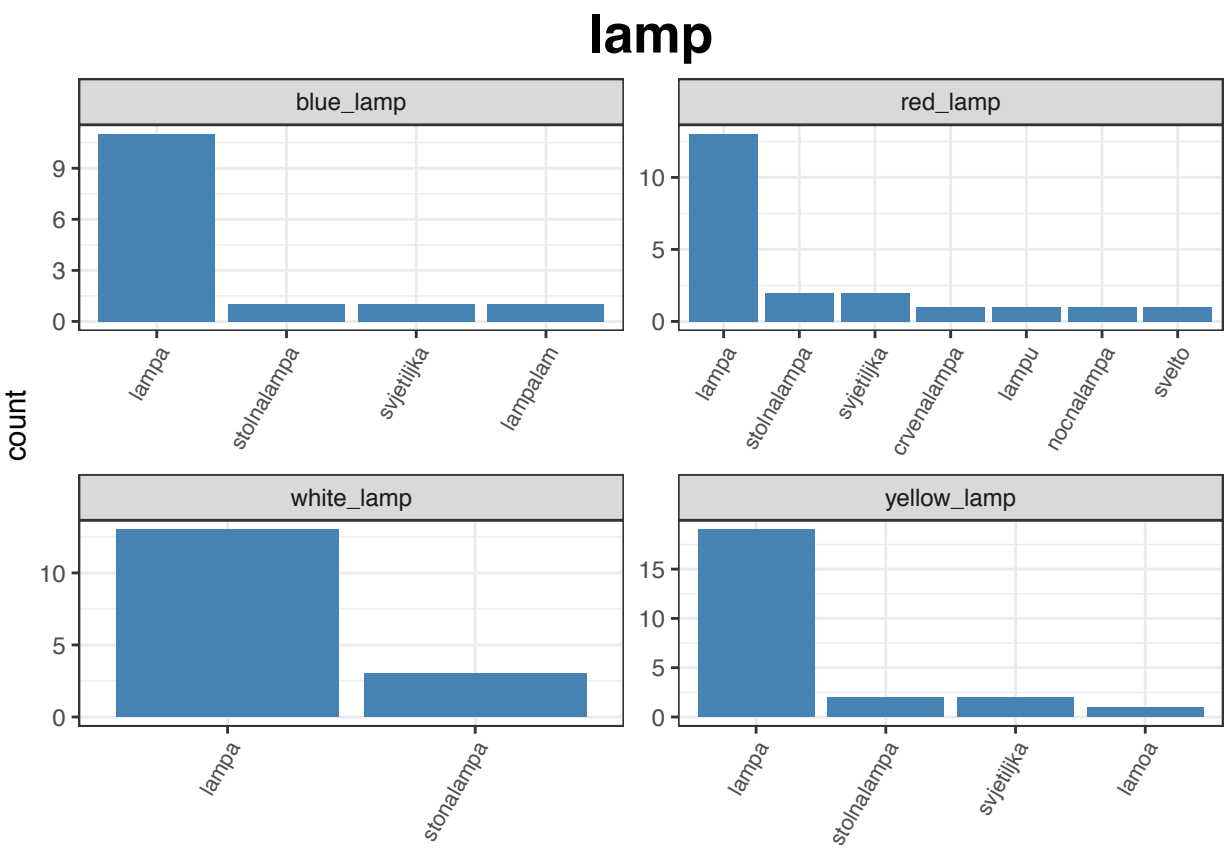
```
##
## [[52]]
```


knife
black_knife
green_knife
orange_knife
purple_knife
count
noz
kuhinjskinoz
crninoz
noz(opetnecerecod3slova)
velikinoz
noz.
nozh
nozz
nozzarezanje
ovojenoz
sjeckalo
zeleninoz
knife
narancastinoz
kuharskinoz
ljubicastnoz
noz...
nozzasecenje

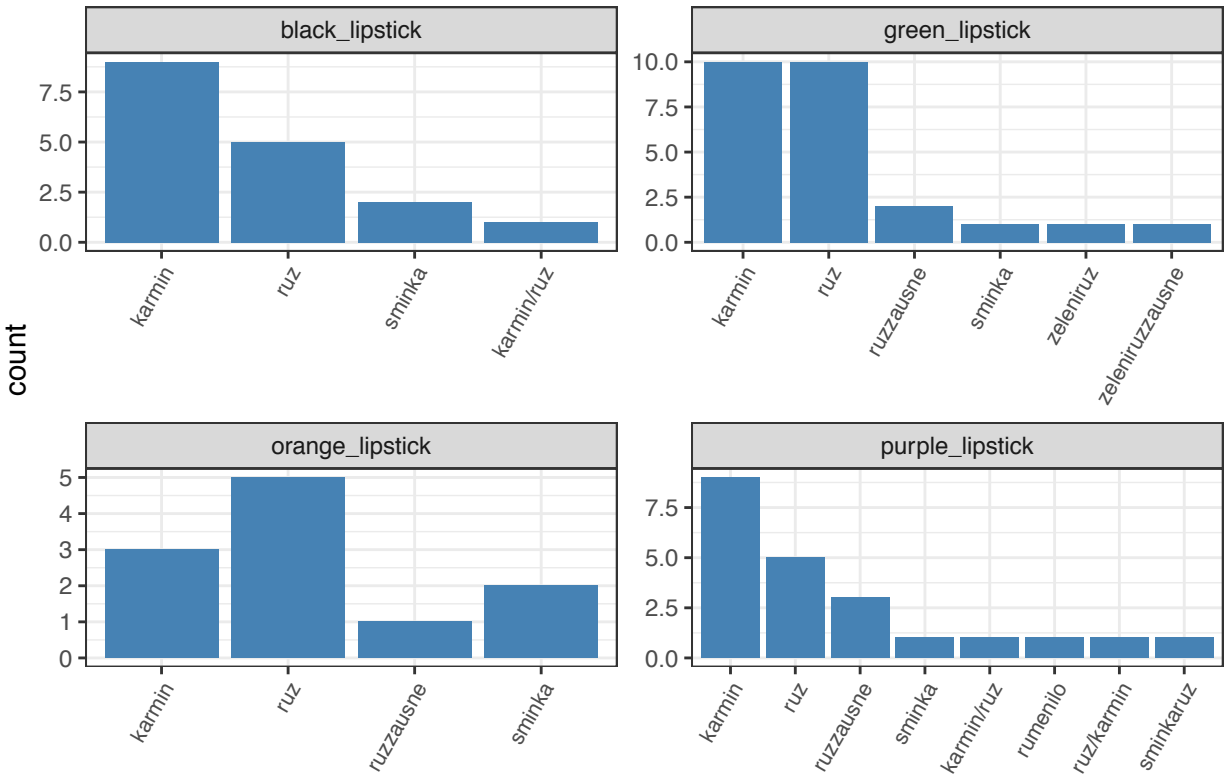
```
##
## [[53]]
```


ladle
black_ladle
green_ladle
orange_ladle
purple_ladle
count
kutlaca
seflja
kuhaca
zaimaca
kaciola
kasika
kuhinjskavelikakasika
sefarka
kutljaca
kacivola
velikazlica
grabilicazaljuhu
ladle
neznam
neznam,onozavaditijuhu
grabilica
kaska
paljaca
sefla
susak

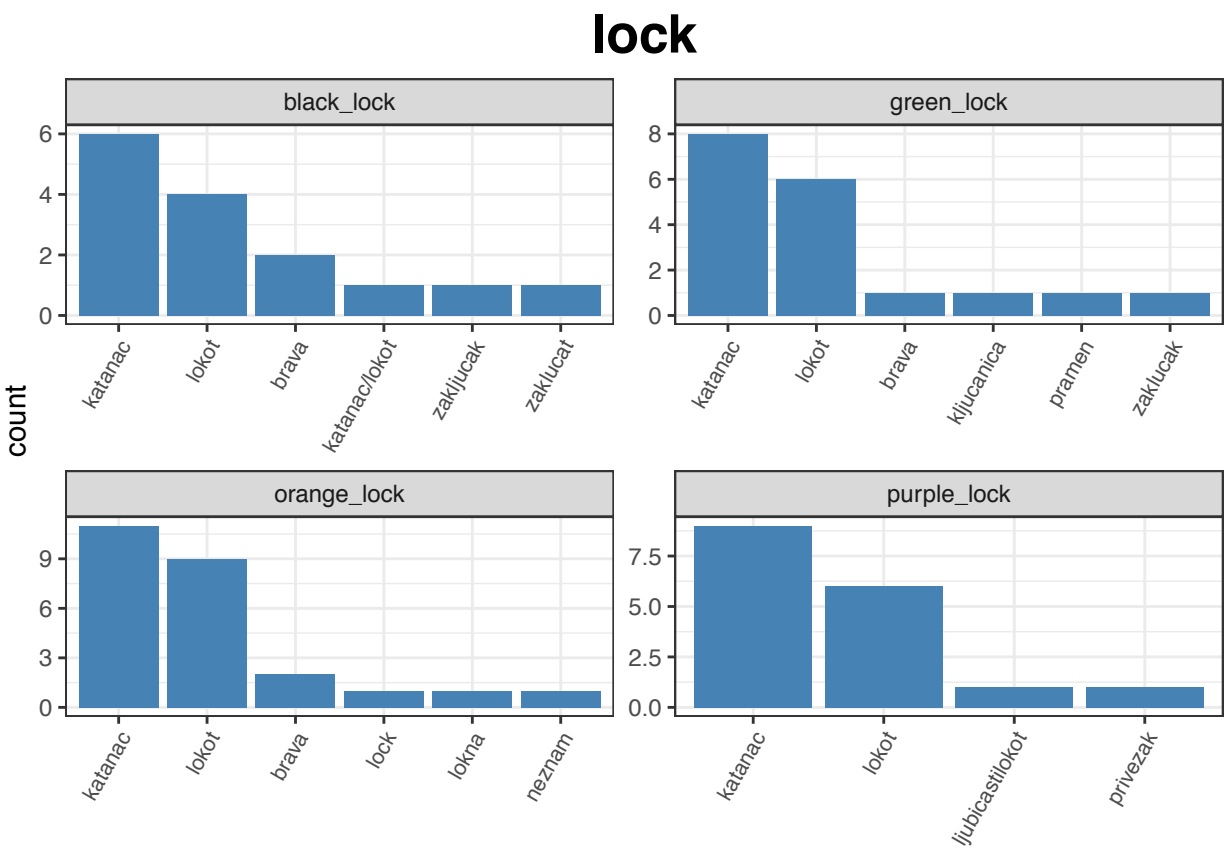
```
##
## [[54]]
```


lamp
blue_lamp
red_lamp
white_lamp
yellow_lamp
count
lampa
stolnalampa
svjetiljka
lampalam
crvenalampa
lampu
nocnalampa
svelto
stonalampa
lamoa

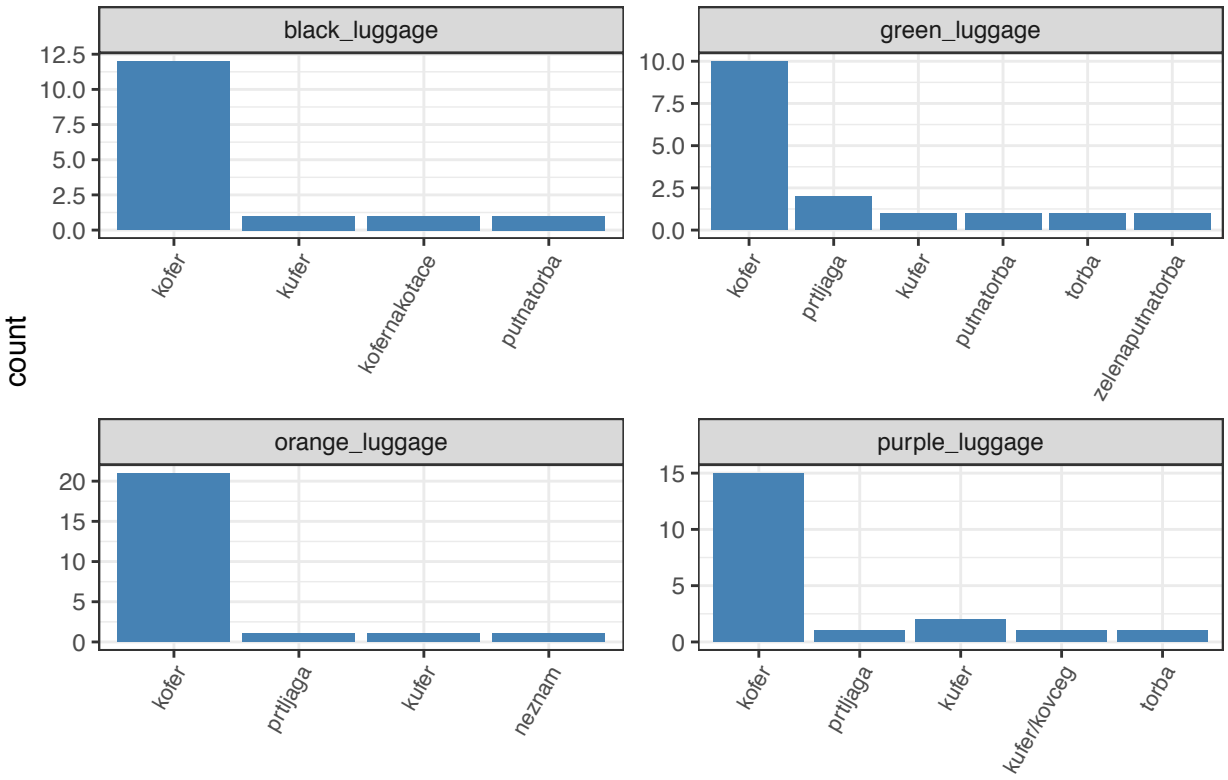
```
##
## [[55]]
```


lipstick
black_lipstick
green_lipstick
orange_lipstick
purple_lipstick
count
karmin
ruz
sminka
karmin/ruz
ruzzausne
zeleniruz
zeleniruzzausne
rumenilo
ruz/karmin
sminkaruz

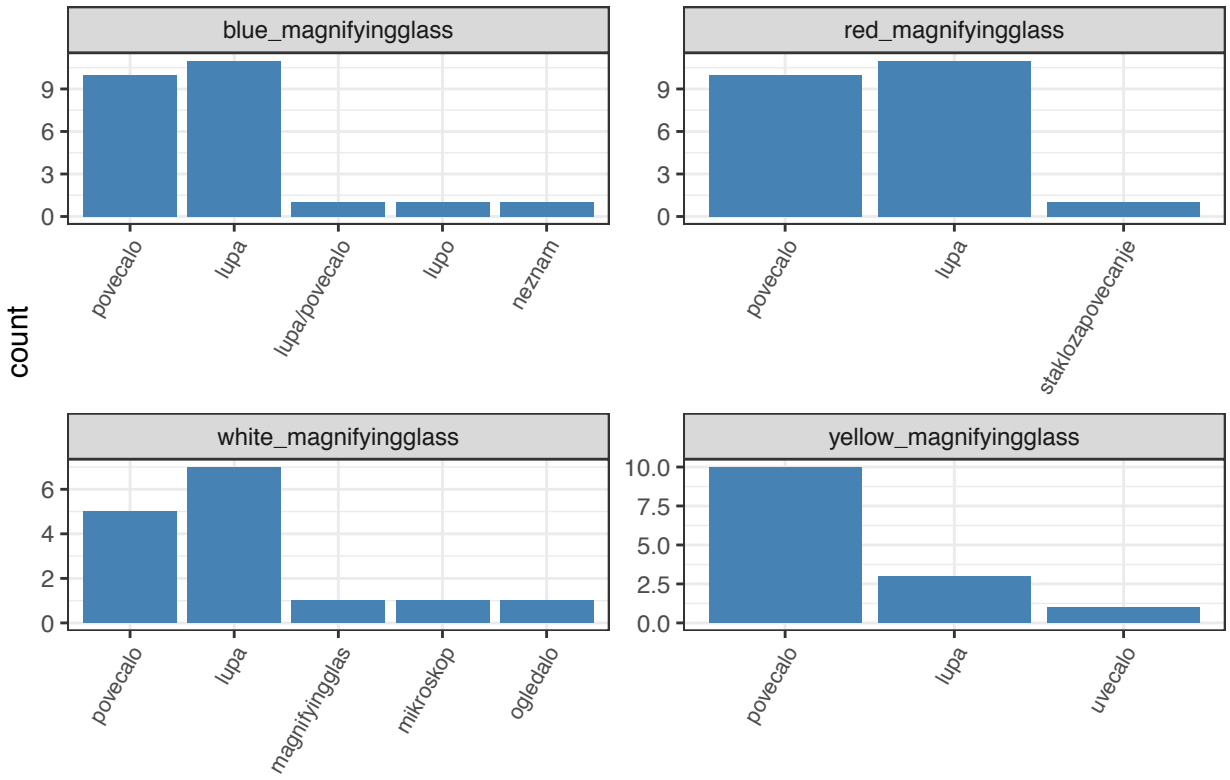
```
##
## [[56]]
```


lock
black_lock
green_lock
orange_lock
purple_lock
count
katanac
lokot
brava
katanac/lokot
zakljucak
zaklucat
kljucanica
pramen
zaklucak
lock
lokna
neznam
ljubicastilokot
privezak

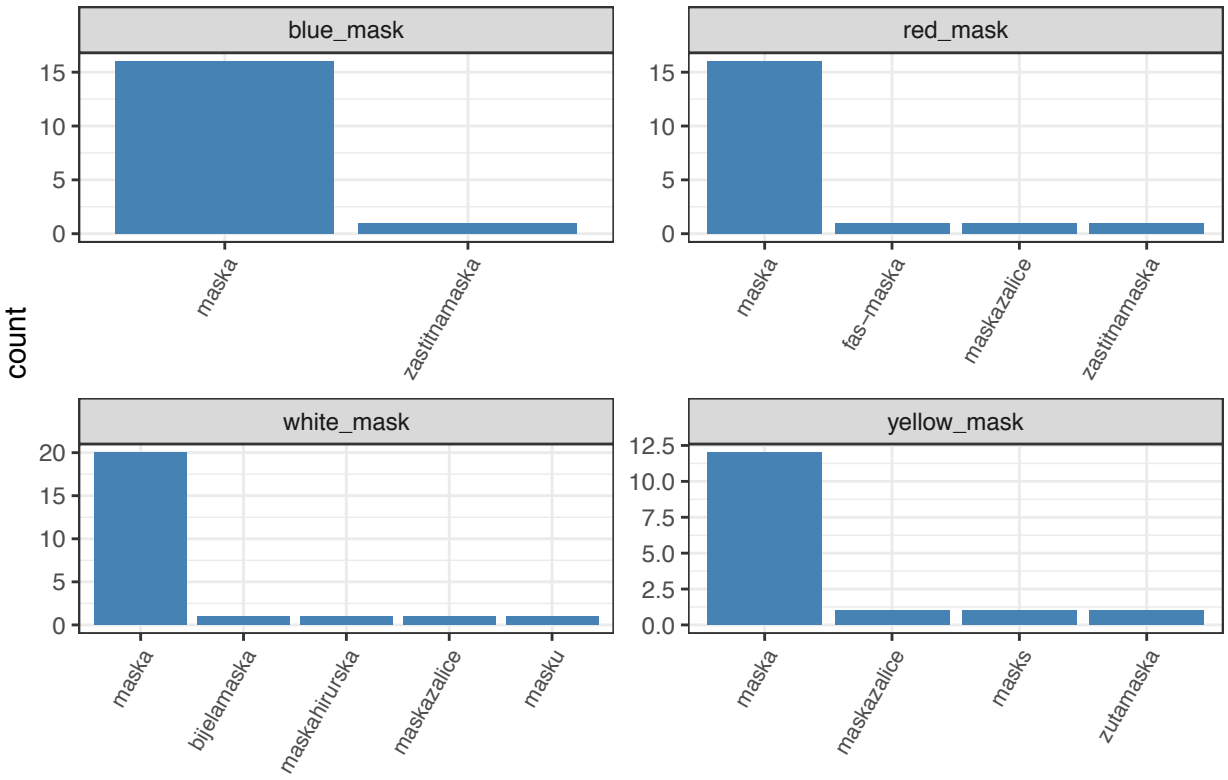
```
##
## [[57]]
```


luggage
black_luggage
green_luggage
orange_luggage
purple_luggage
count
kofer
kufer
koffernakotace
putnatorba
prtljaga
putnatorba
torba
zelenaputnatorba
neznam
kufer/kovceg

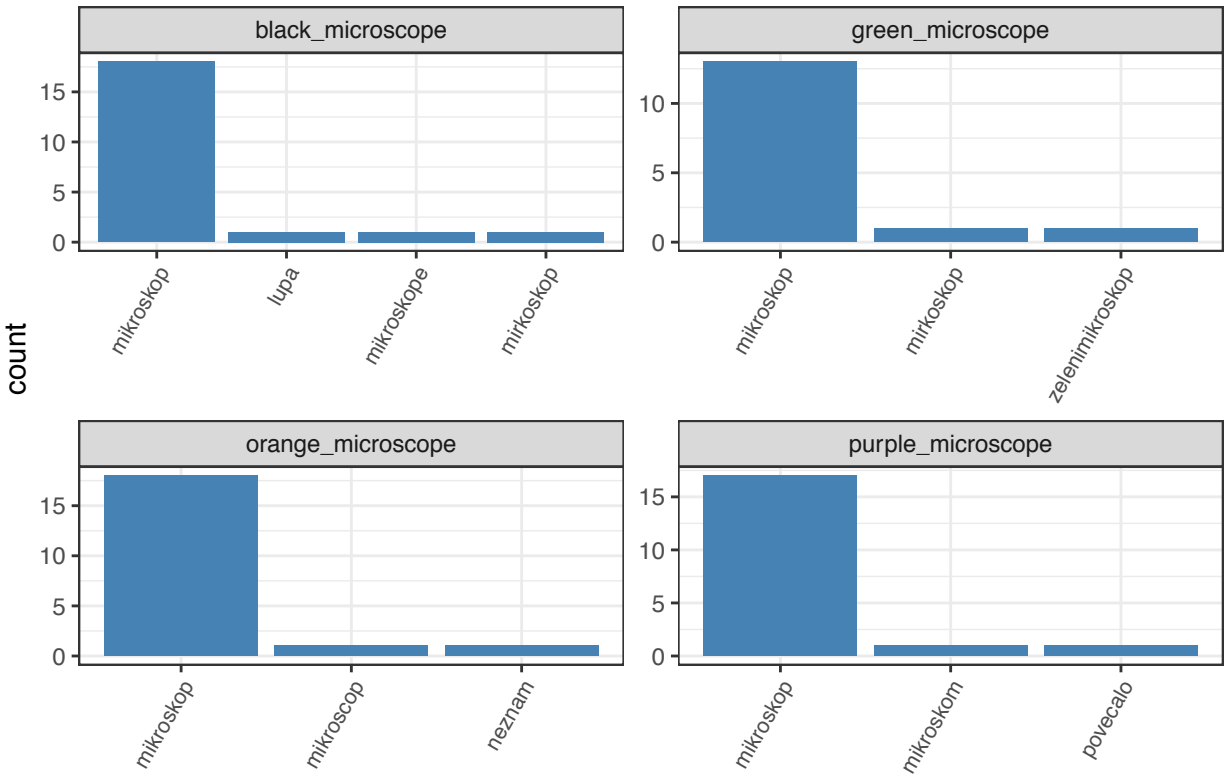
```
##
## [[58]]
```


magnifyingglass
blue_magnifyingglass
red_magnifyingglass
white_magnifyingglass
yellow_magnifyingglass
count
povecalo
lupa
lupa/povecalo
lupo
neznam
staklozapovecanje
magnifyingglas
mikroskop
ogledalo
uvecalo

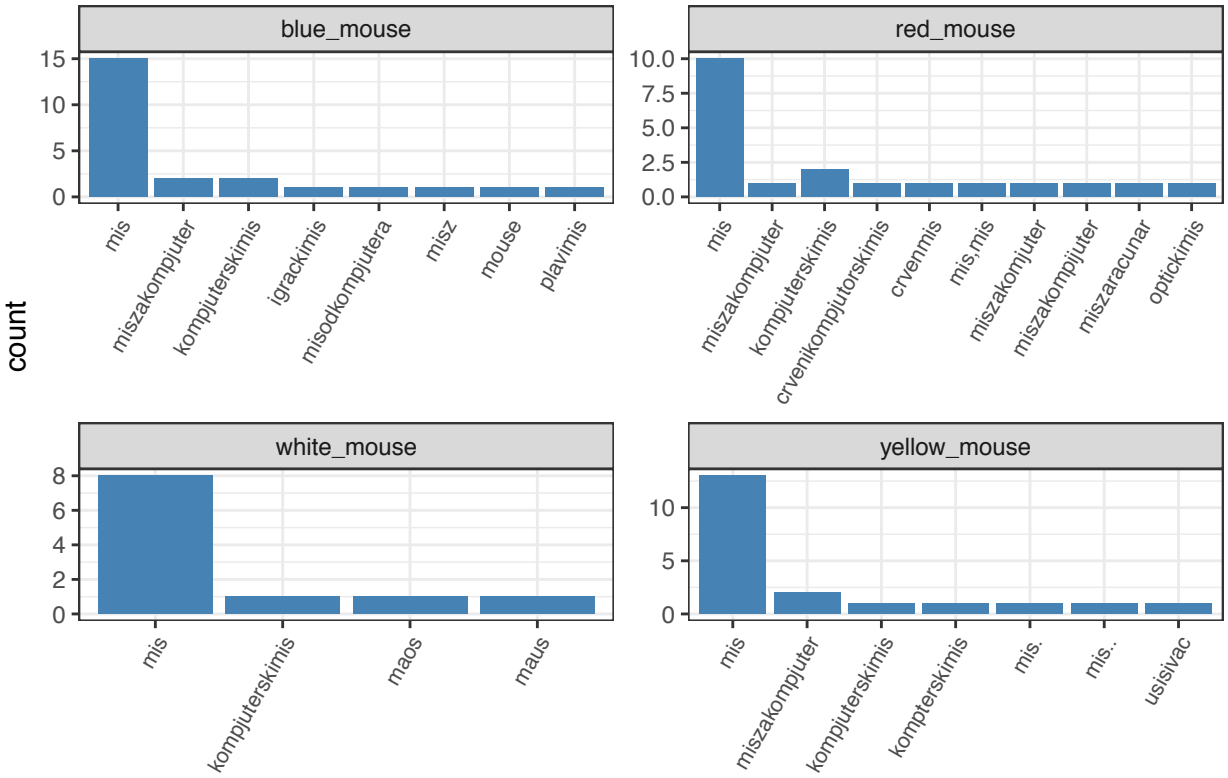
```
##
## [[59]]
```


mask
blue_mask
red_mask
white_mask
yellow_mask
count
maska
zastitnamaska
fas-maska
maskazalice
bijelamaska
maskahirurska
masku
masks
zutamaska

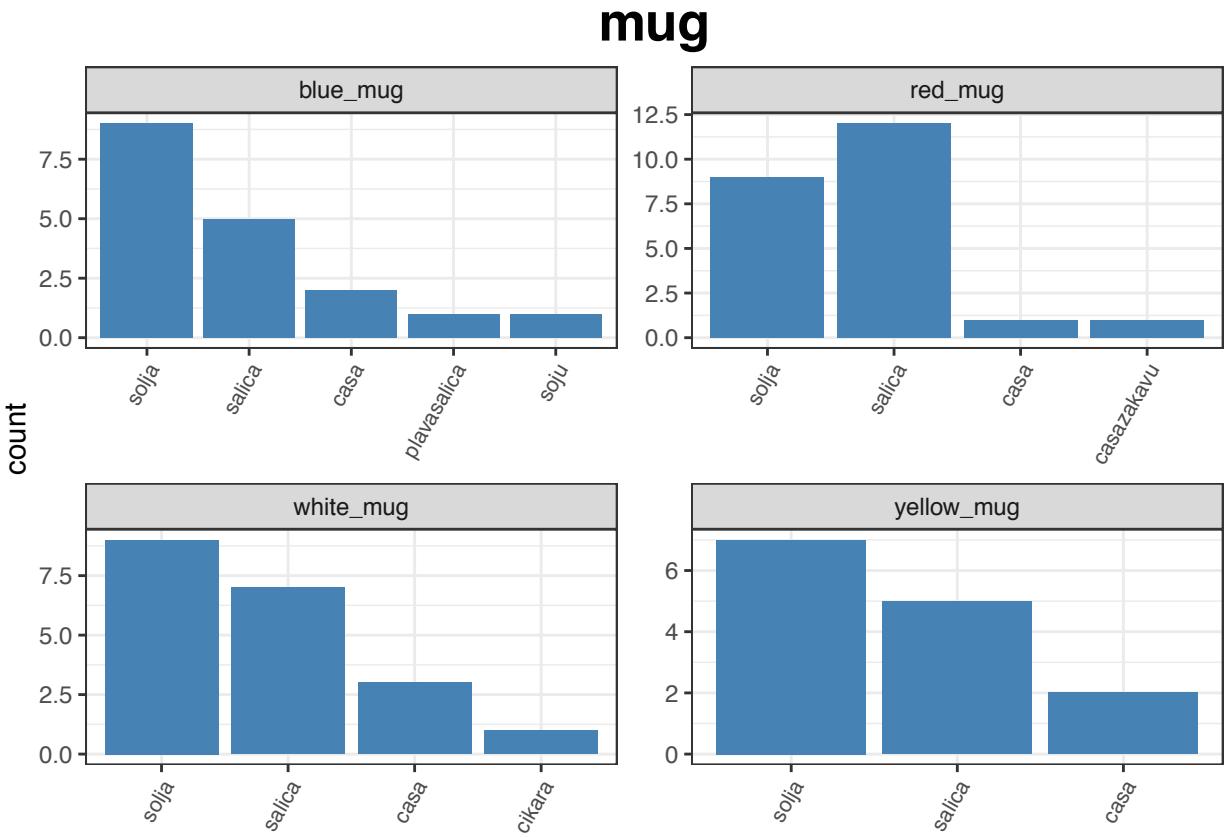
```
##
## [[60]]
```


microscope
black_microscope
green_microscope
orange_microscope
purple_microscope
count
mikroskop
lupa
mikroskope
mirkoskop
mikroskop
mirkoskop
zelenimikroskop
mikroskop
mikroscop
neznam
mikroskop
mikroskom
povecalo

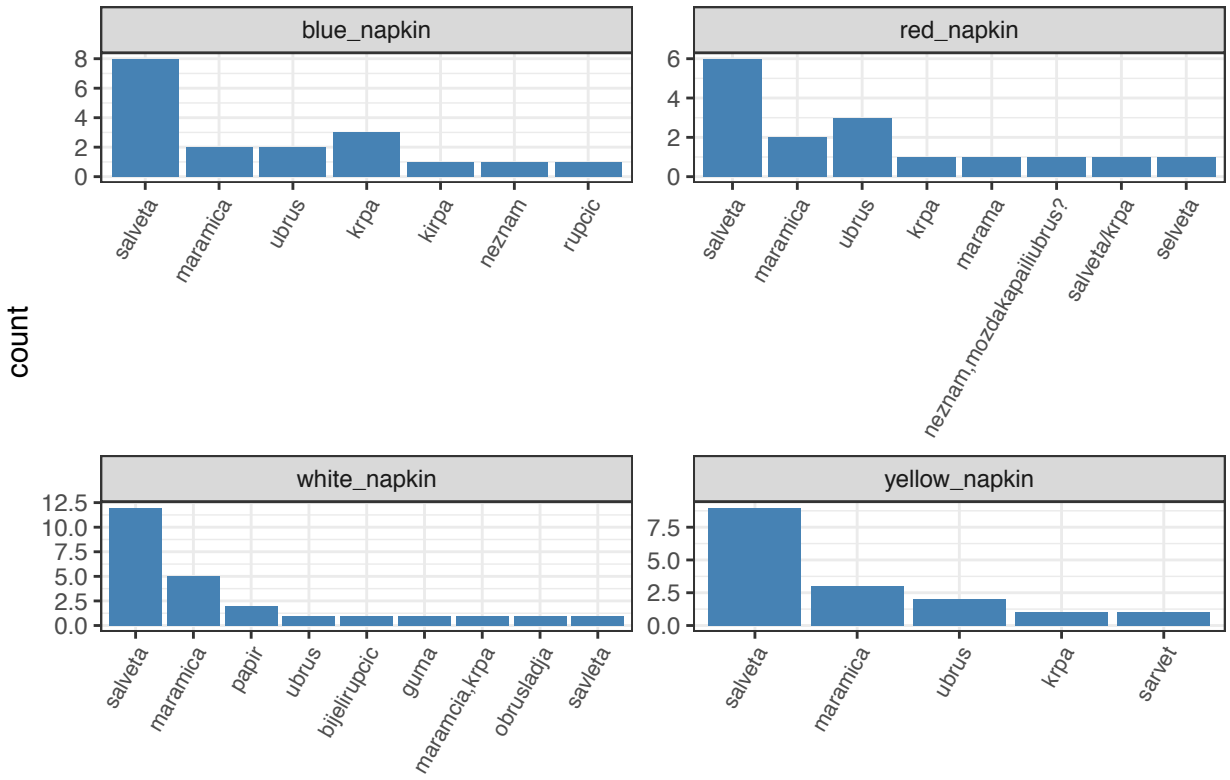
```
##
## [[61]]
```


mouse
blue_mouse
red_mouse
white_mouse
yellow_mouse
count
mis
miszakompjuter
kompjuterskimis
igrackimis
misodkompjutera
misz
mouse
playimis
crvenikompjutorskimis
crvenmis
mis,mis
miszakomjuter
miszakompjjuter
miszaracunar
optickimis
kompjuterskimis
maos
maus
kompterskimis
mis.
mis..
usisivac

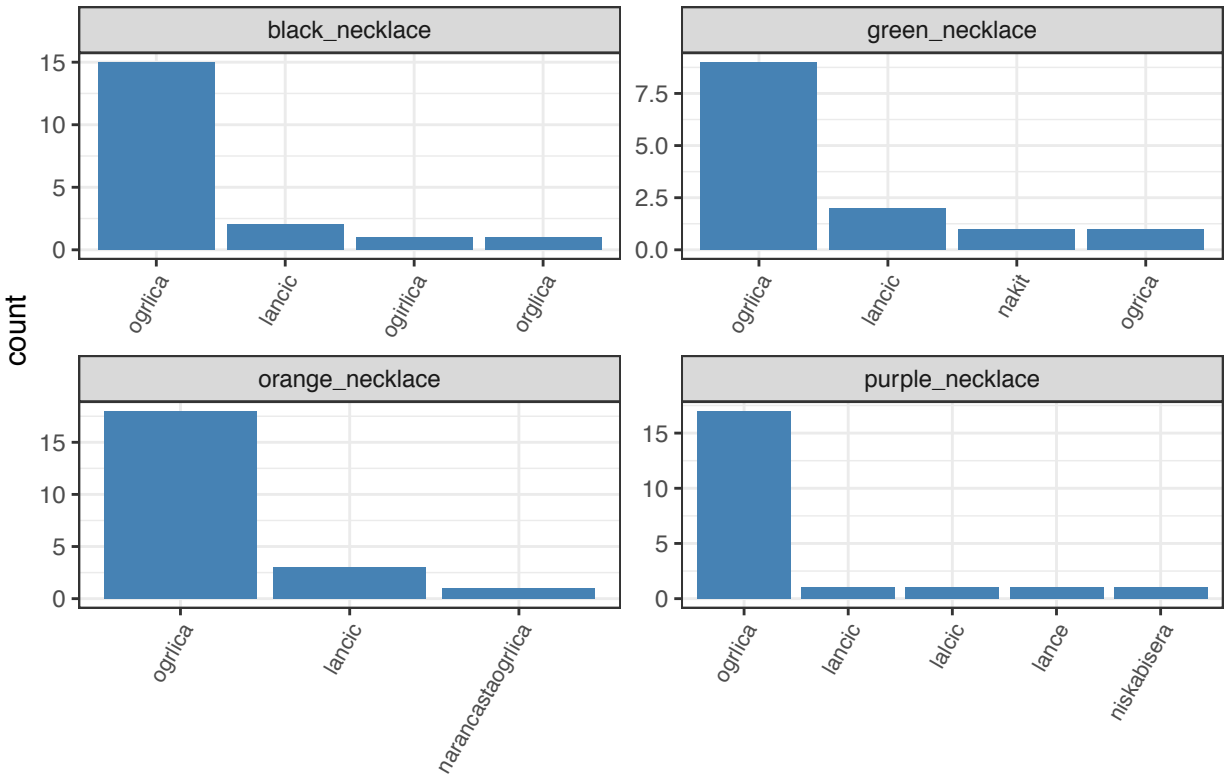
```
##
## [[62]]
```


mug
blue_mug
red_mug
white_mug
yellow_mug
count
solja
salica
casa
plavasalica
soju
casazakavu
cikara

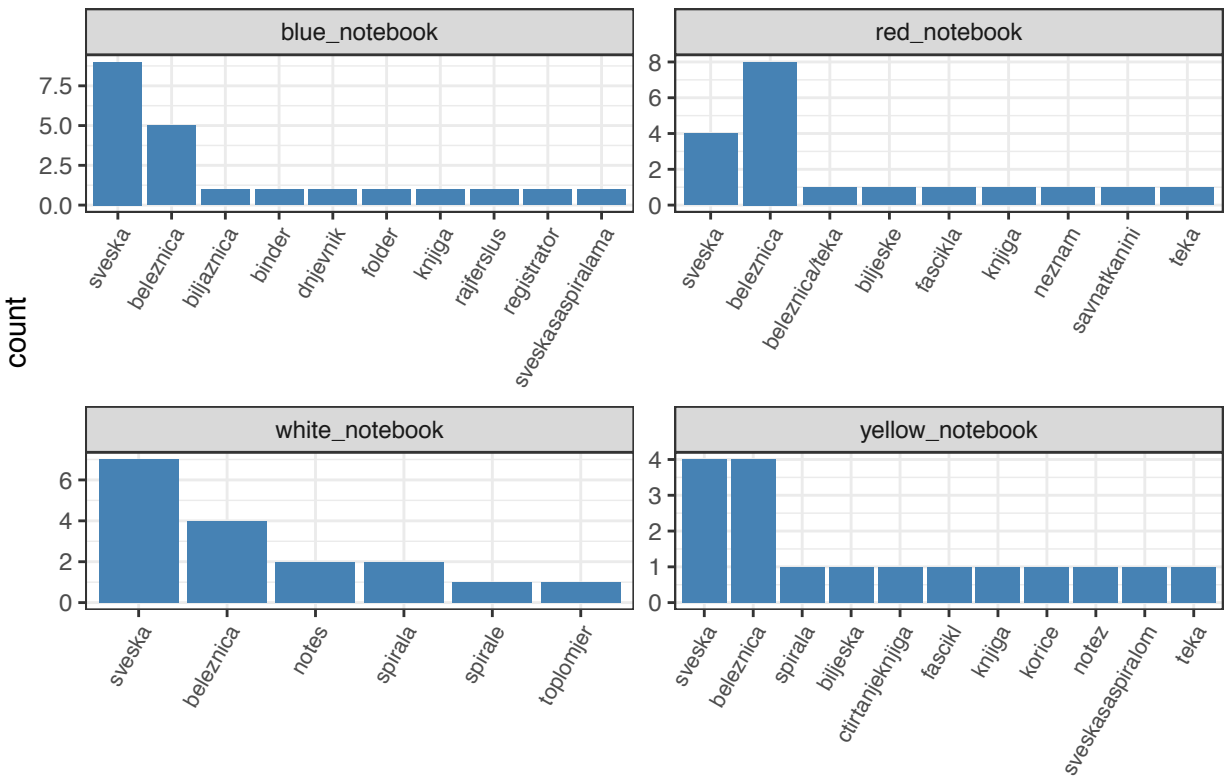
```
##
## [[63]]
```


napkin
blue_napkin
red_napkin
white_napkin
yellow_napkin
count
salveta
maramica
ubrus
krpa
kirpa
neznam
rupcic
marama
neznam,mozdakapalilubrus?
salveta/krpa
selveta
papir
bijelirupcic
guma
maramcia,krpa
obrusladja
savleta
sarvet

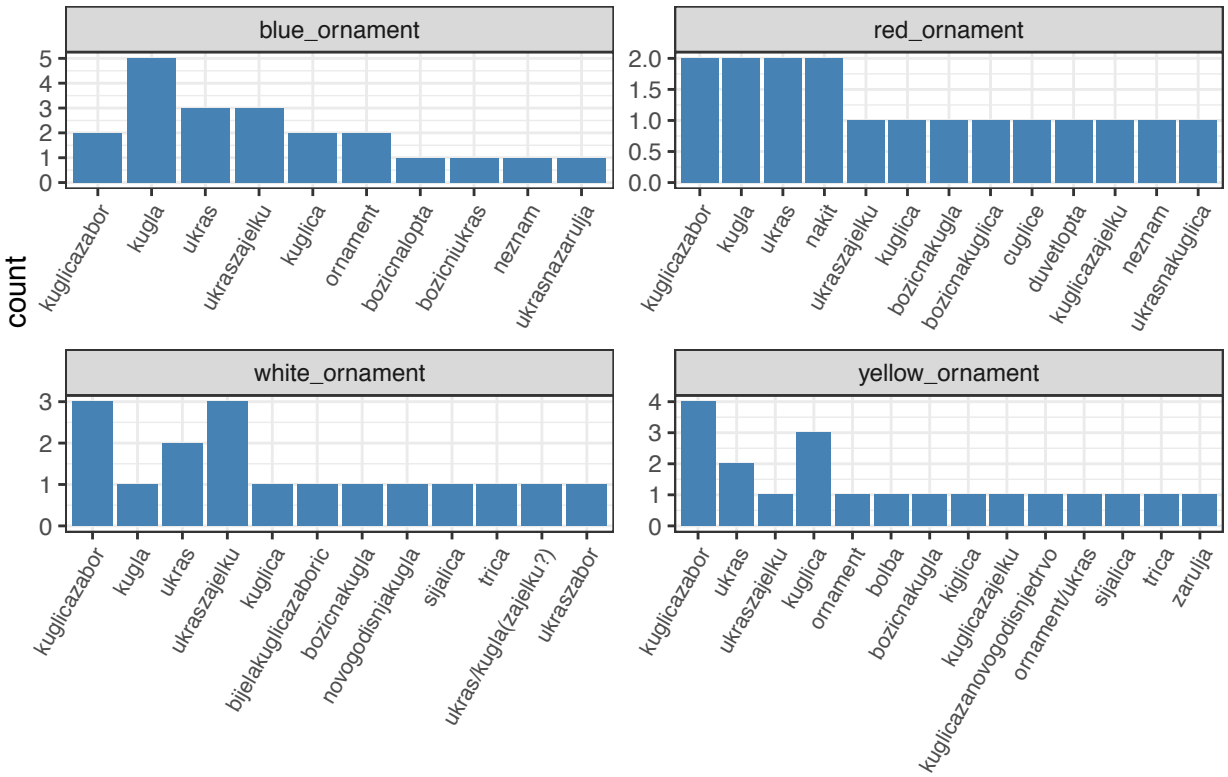
```
##
## [[64]]
```


necklace
black_necklace
green_necklace
orange_necklace
purple_necklace
count
ogrlica
lancic
ogirlica
orglica
nakit
ogrica
narancastaogrlica
lalcic
lance
niskabisera

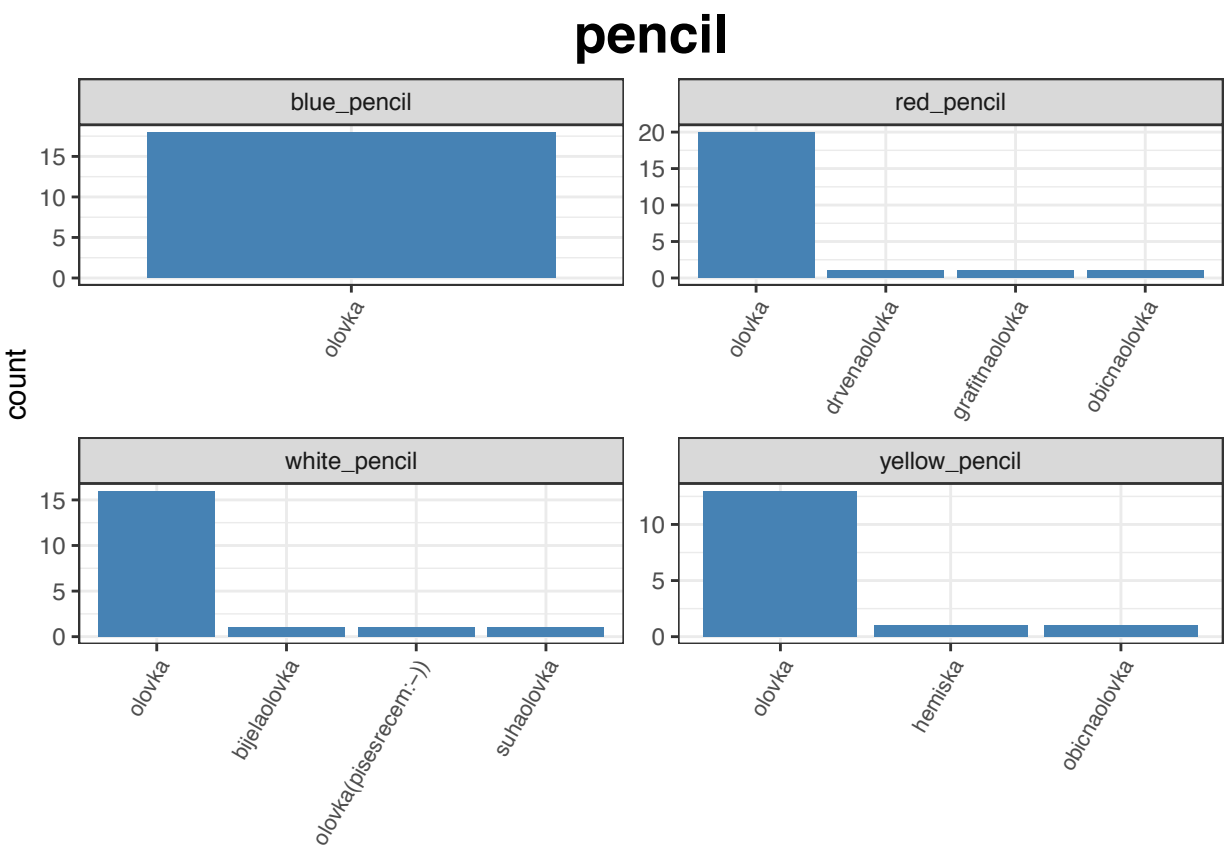
```
##
## [[65]]
```


notebook
blue_notebook
red_notebook
white_notebook
yellow_notebook
count
sveska
beleznica
biljaznica
binder
dnjevnik
folder
knjiga
rajferslus
registrator
sveskasaspiralama
beleznica/teka
bilješke
fascikla
neznam
savnatkanini
teka
notes
spirala
spirale
toplomjer
bilješka
ctirtanjeknjiga
fascikl
knjiga
korice
notez
sveskasaspiralom

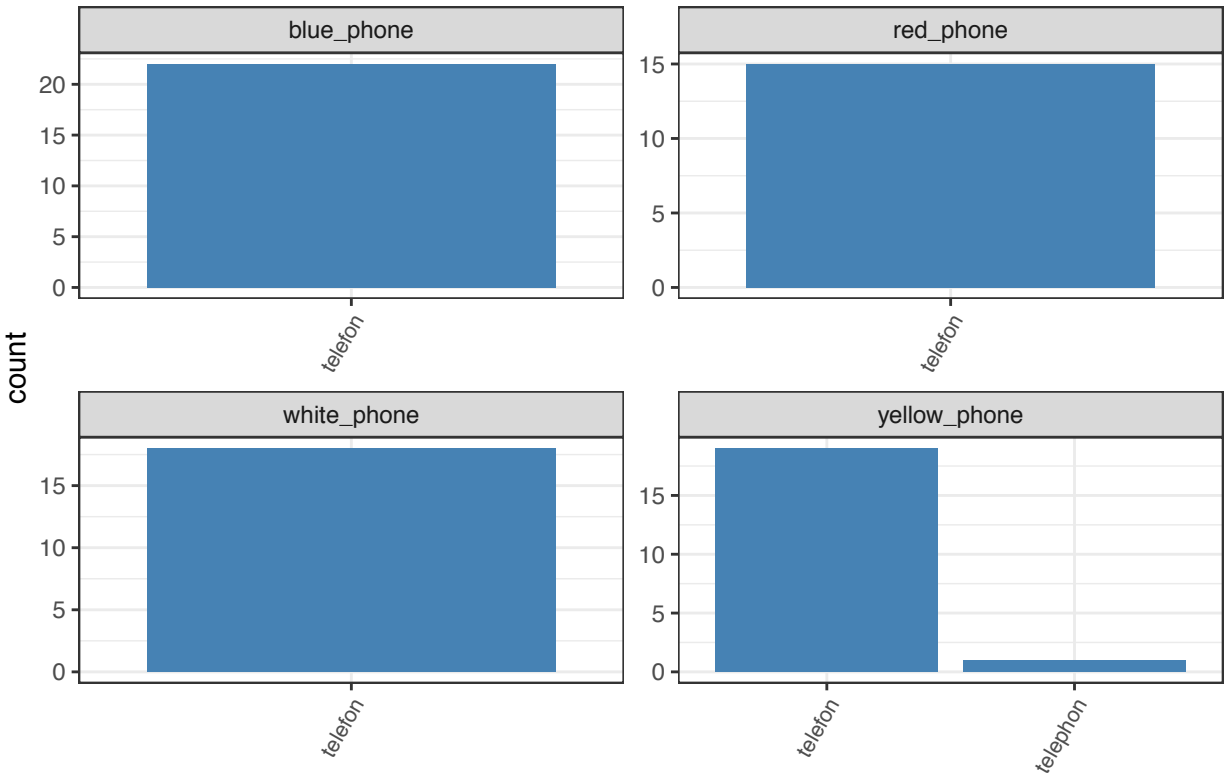
```
##
## [[66]]
```


ornament
blue_ornament
red_ornament
white_ornament
yellow_ornament
count
kuglicazabor
kugla
ukras
ukraszajelku
kuglica
ornament
bozicnalopta
bozicniukras
neznam
ukrasnazarulja
nakit
bozicnakugla
bozicnakuglica
cuglice
duvetlopta
kuglicazajelku
ukrasnakuglica
bijelakuglicazaboric
novogodisnjakugla
sijalica
trica
ukras/kugla(zajelku?)
ukraszabor
bolba
kiglica
kuglicazajelku
kuglicazanovogodisnjedrvo
ornament/ukras
zarulja

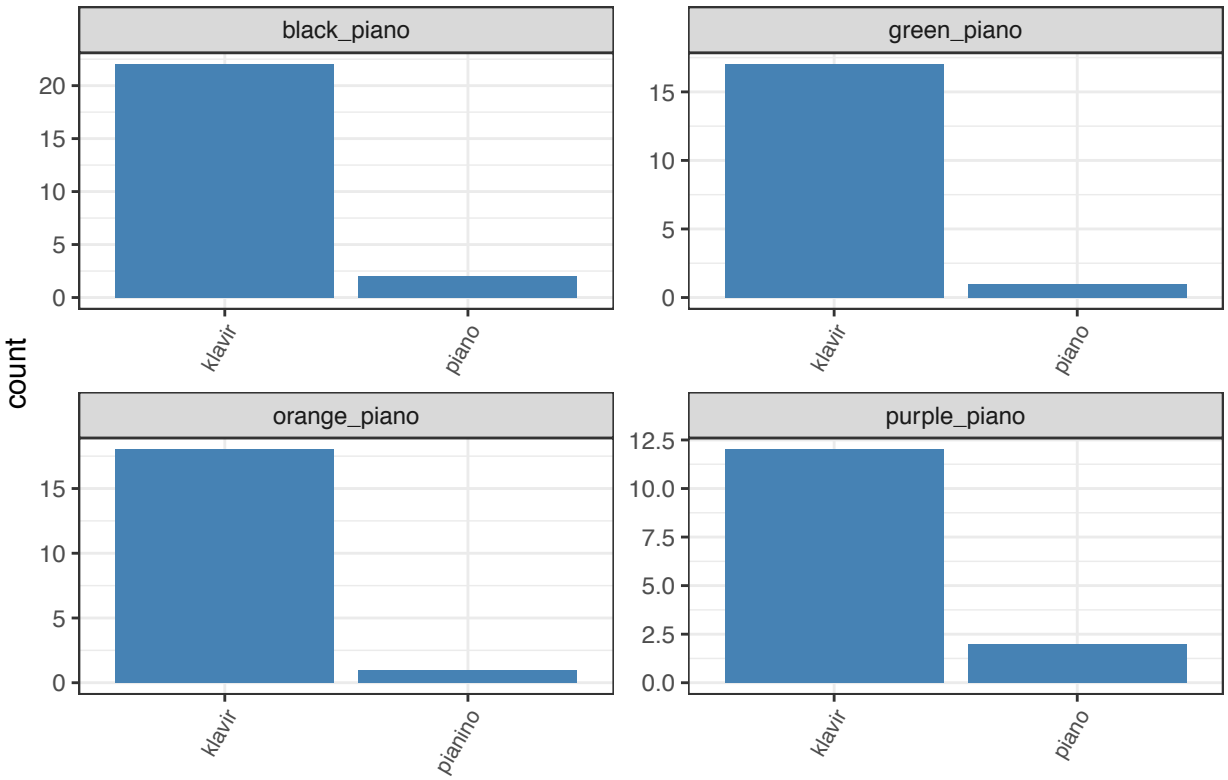
```
##
## [[67]]
```


pencil
blue_pencil
red_pencil
white_pencil
yellow_pencil
count
olovka
olovka
drvenaolovka
grafitnaolovka
obicnaolovka
olovka
bijelaolovka
olovka(pisesrecem:-))
suhaolovka
olovka
hemiska
obicnaolovka

```
##
## [[68]]
```


phone
blue_phone
red_phone
white_phone
yellow_phone
count
telefon
telephon

```
##
## [[69]]
```


piano
black_piano
green_piano
orange_piano
purple_piano
count
20
15
10
5
0
12.5
10.0
7.5
5.0
2.5
0.0
klavir
piano
pianino

```
##
## [[70]]
```

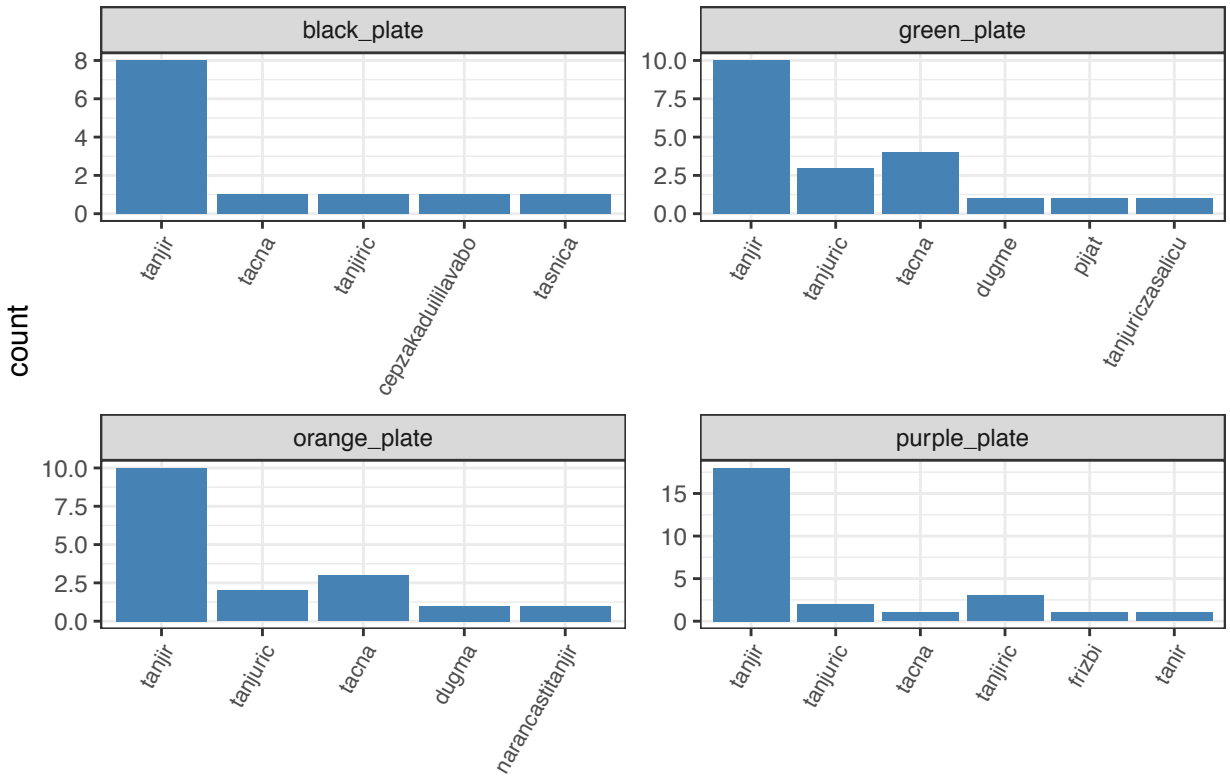
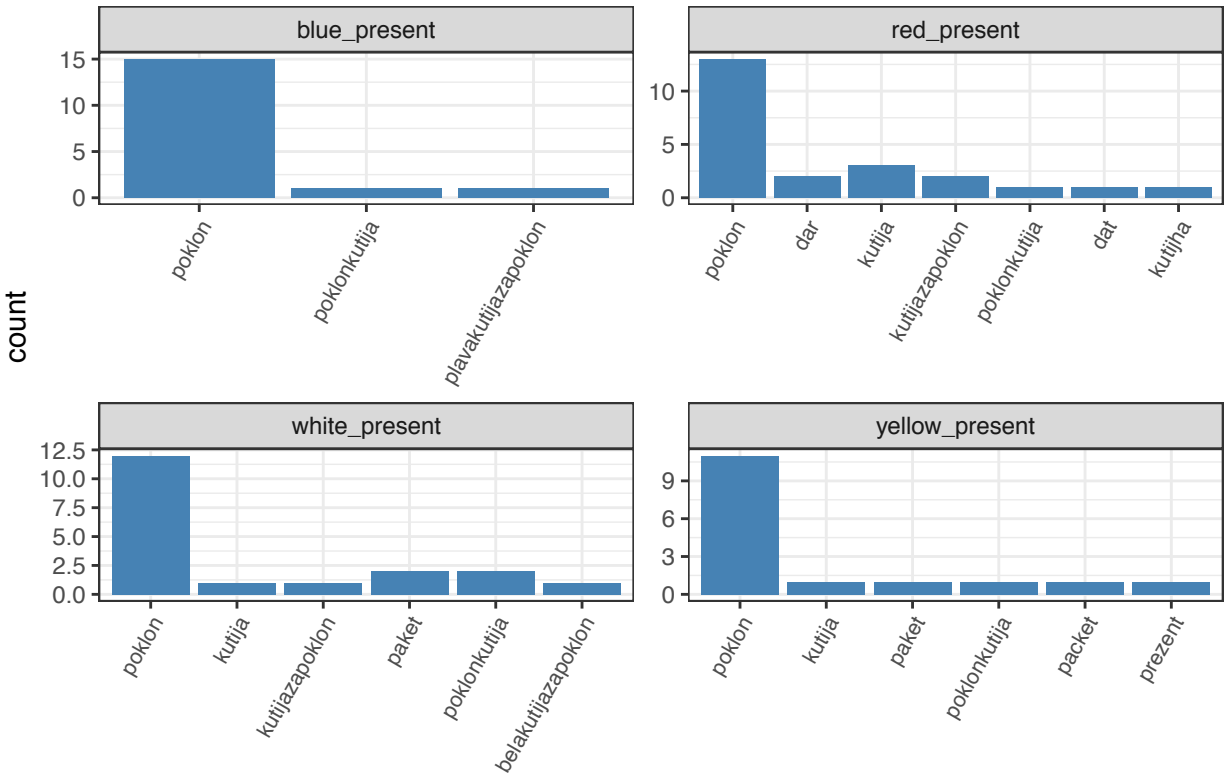
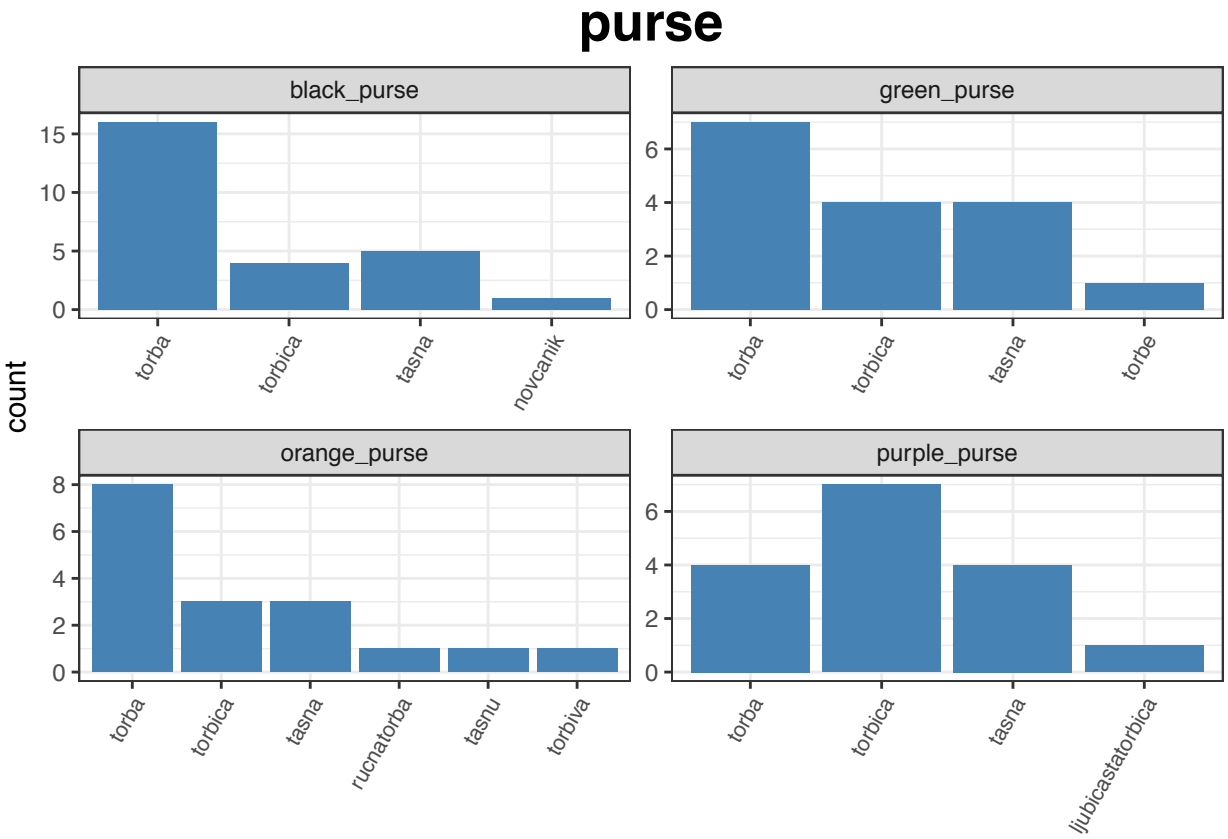


plate
black_plate
green_plate
orange_plate
purple_plate
count
tanjir
tacna
tanjiric
cepzakaduililavabo
tasnica
tanjuric
dugme
pijat
tanjuriczasalicu
dugma
narancastitanjir
frizbi
tanir

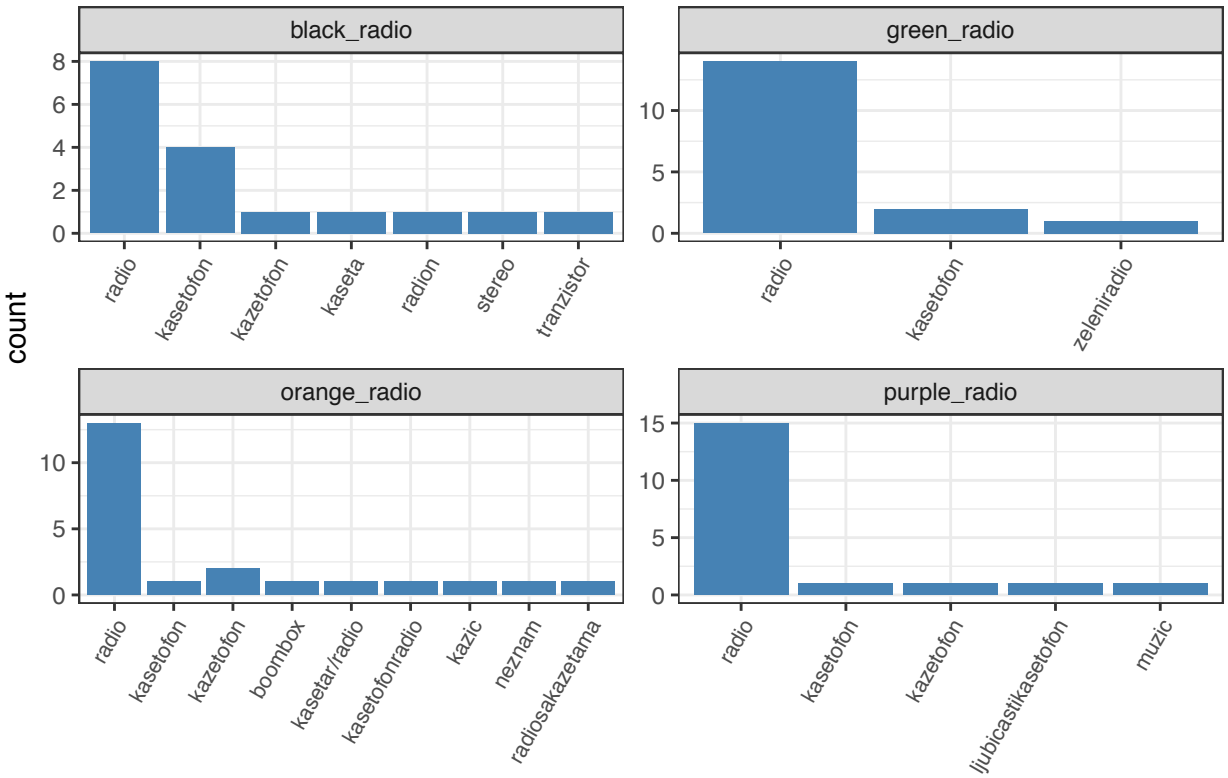
```
##
## [[71]]
```


present
blue_present
red_present
white_present
yellow_present
count
poklon
poklonkutija
plavakutijazapoklon
dar
kutija
kutijazapoklon
poklonkutija
dat
kutijha
kutija
kutijazapoklon
paket
poklonkutija
belakutijazapoklon
packet
prezent

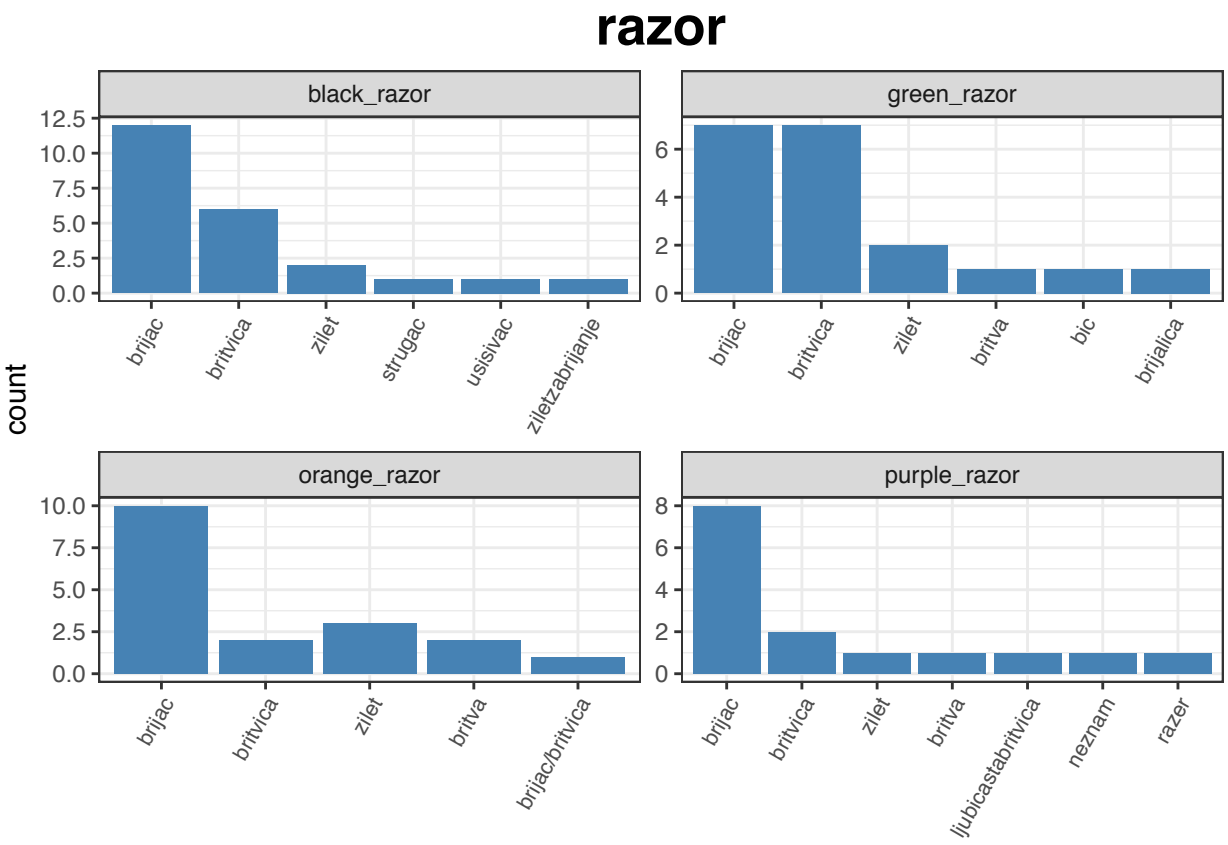
```
##
## [[72]]
```


purse
black_purse
green_purse
orange_purse
purple_purse
count
torba
torbica
tasna
novcanik
torbe
rucnatorba
tasnu
torbiva
ljubicastatorbica

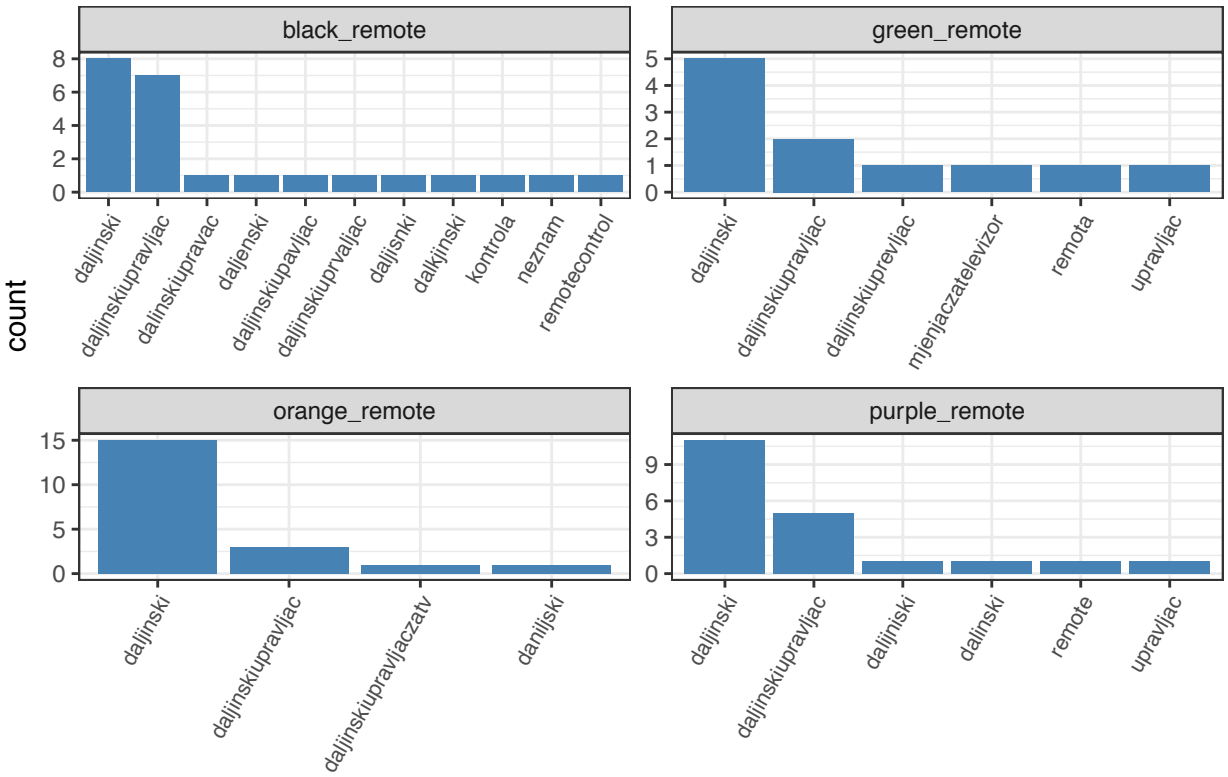
```
##
## [[73]]
```


radio
black_radio
green_radio
orange_radio
purple_radio
count
radio
kasetofon
kazetofon
kaseta
radion
stereo
tranzistor
zeleniradio
boombox
kasetar/radio
kasetofonradio
kazic
neznam
radiosakazetama
ljubicastikasetofon
muzic

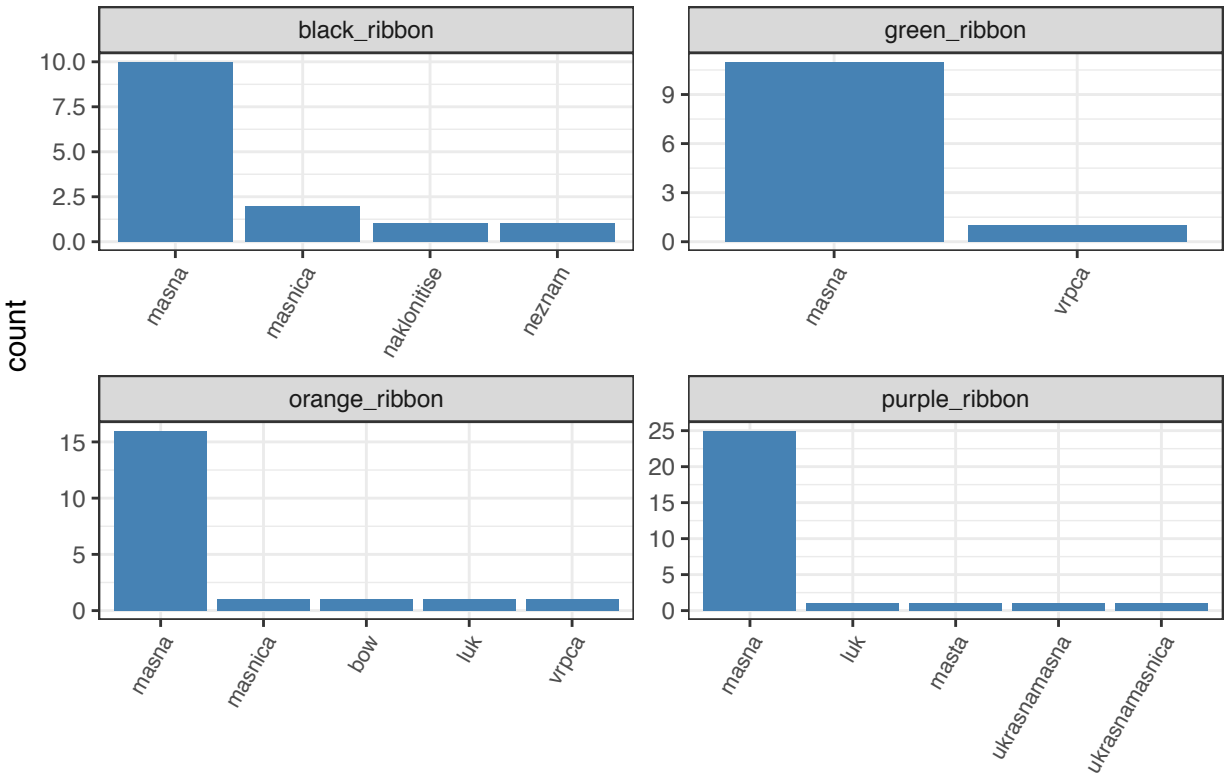
```
##
## [[74]]
```


razor
black_razor
green_razor
orange_razor
purple_razor
count
brijac
britvica
zilet
strugac
usisivac
ziletzabrijanje
britva
bic
brijalica
brijac/britvica
ljubicastabritvica
neznam
razer

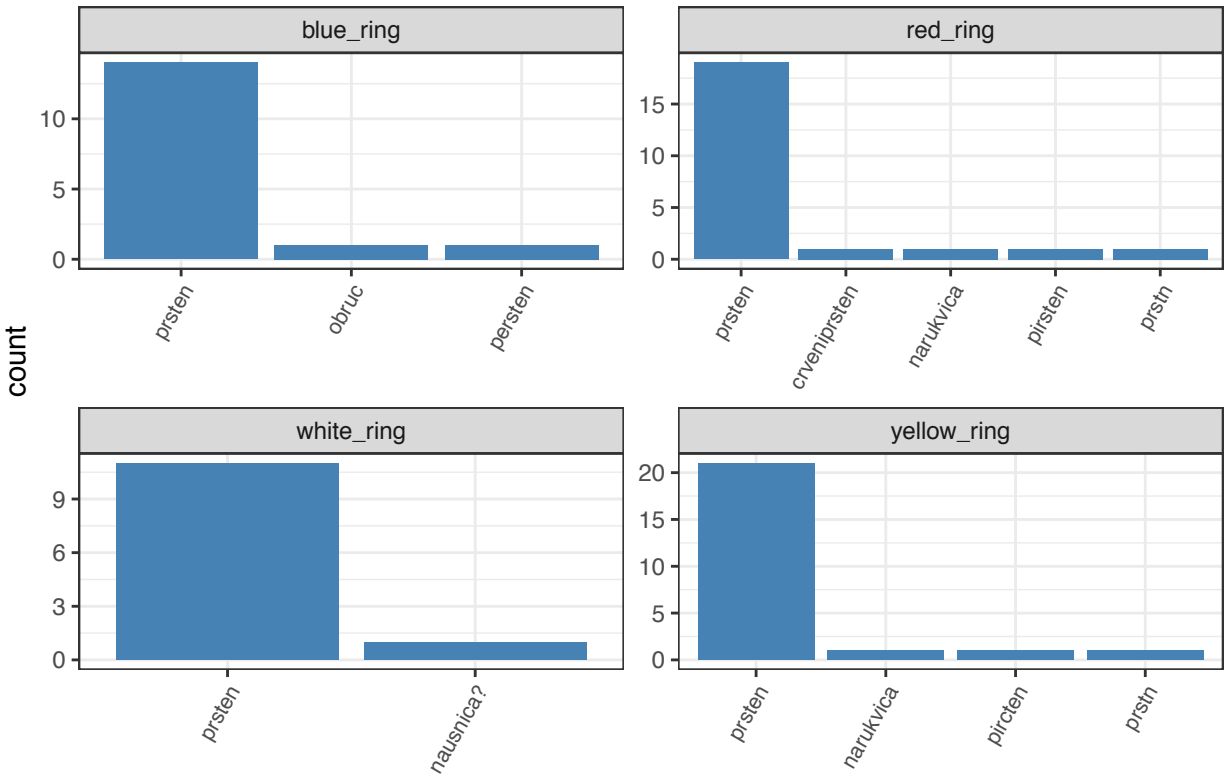
```
##
## [[75]]
```


remote
black_remote
green_remote
orange_remote
purple_remote
count
daljinski
daljinskiupravljac
dalinskiupravac
daljenski
daljinskiupavljac
daljinskiuprvaljac
daljisnki
dalkjinski
kontrola
neznam
remotecontrol
daljinski
daljinskiupravljac
daljinskiuprevljac
mjenjaczatelevizor
remota
upravljac
daljinski
daljinskiupravljac
daljinskiupravljaczatv
daniljski
daljinski
daljinskiupravljac
daljinjski
dalinski
remote
upravljac

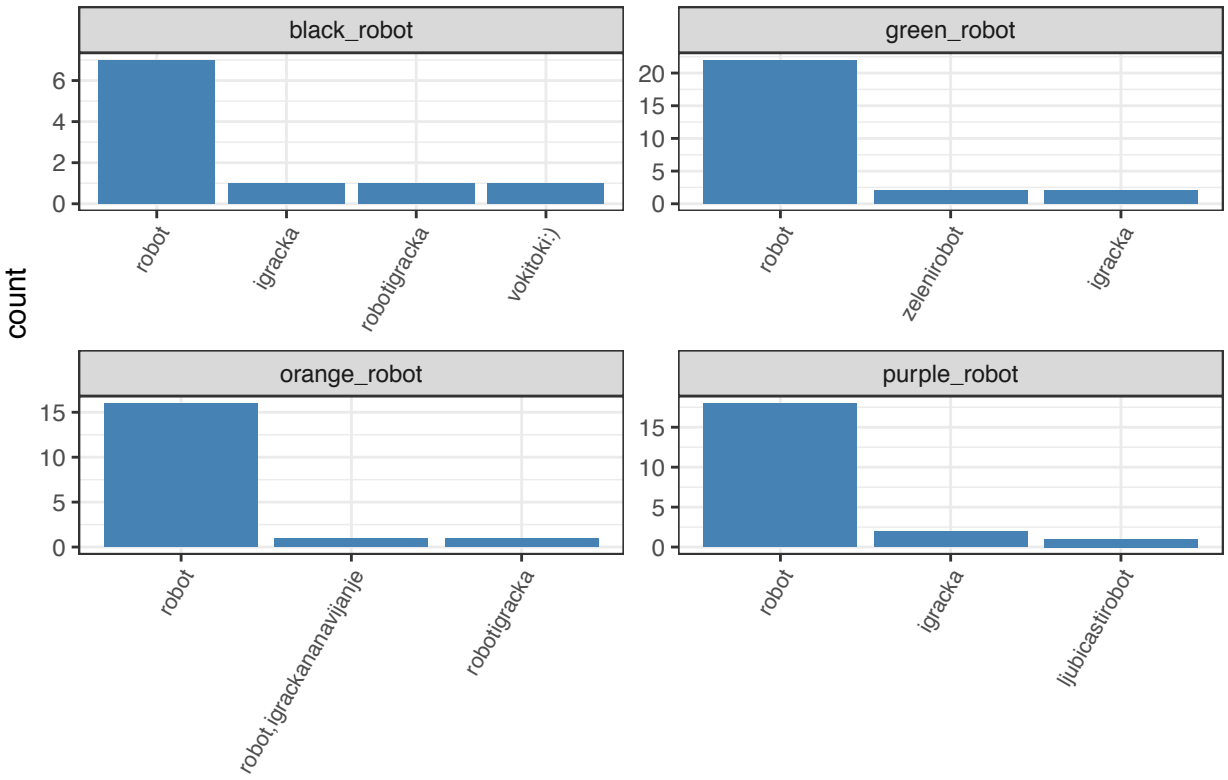
```
##
## [[76]]
```


ribbon
black_ribbon
green_ribbon
orange_ribbon
purple_ribbon
count
masna
masnica
naklonitise
neznam
vrpca
bow
luk
masta
ukrasnamasna
ukrasnamasnica

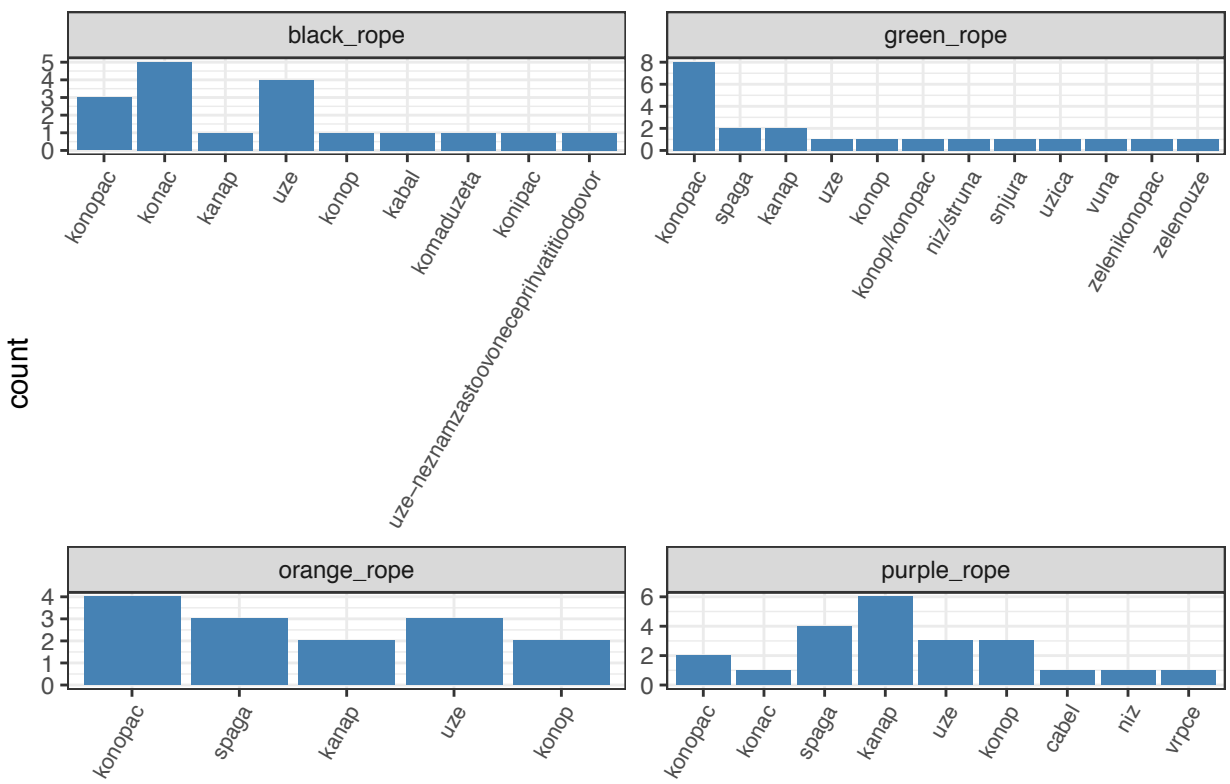
```
##
## [[77]]
```


ring
blue_ring
red_ring
white_ring
yellow_ring
count
prsten
obruc
persten
crveniprsten
narukvica
pirsten
prstn
nausnica?
pircten

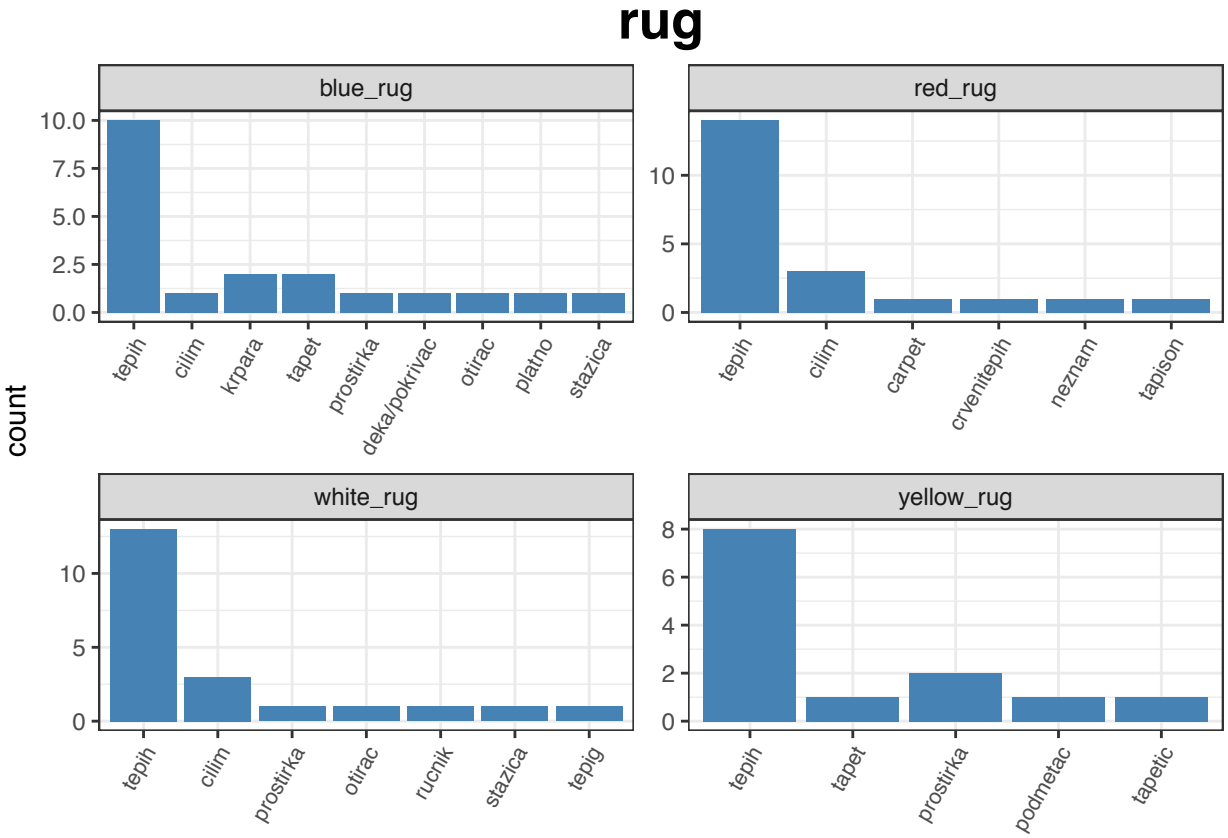
```
##
## [[78]]
```


robot
black_robot
green_robot
orange_robot
purple_robot
count
robot
igracka
robotigracka
vokitoki:)
zelenirobot
robot,igrackananavijanje
ljubicastirobot

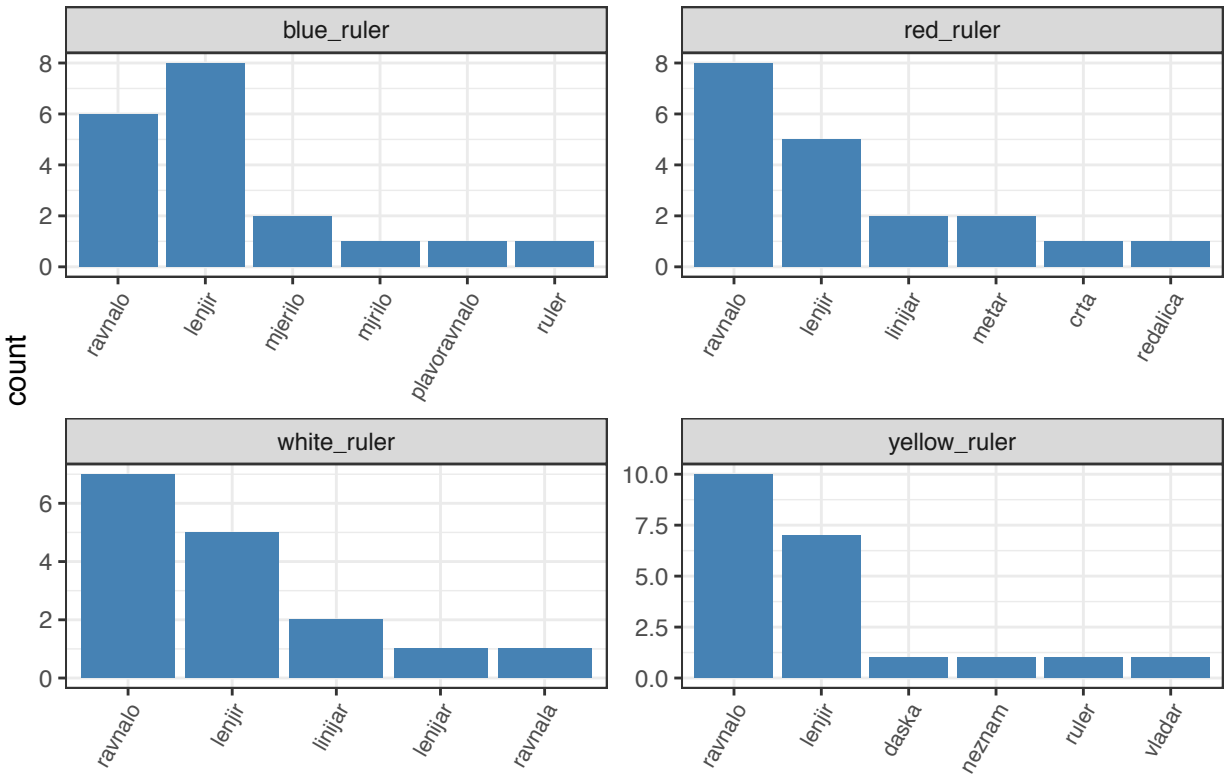
```
##
## [[79]]
```


rope
black_rope
green_rope
orange_rope
purple_rope
count
konopac
konac
kanap
uze
konop
kabal
komaduzeta
konipac
uze-neznamzastoovoneceprihvatitiodgovor
spaga
konop/konopac
niz/struna
snjura
uzica
vuna
zelenikonopac
zelenouze
cabel
niz
vrpce

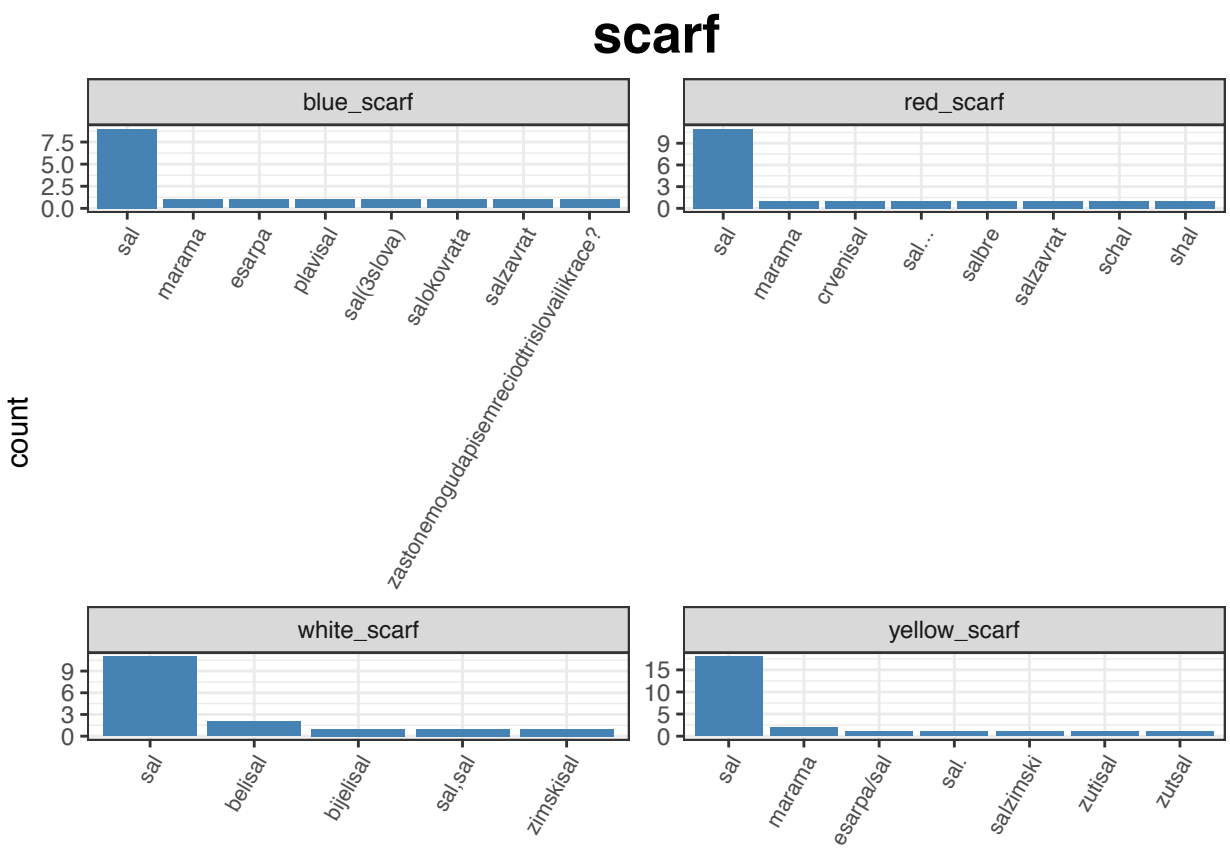
```
##
## [[80]]
```


rug
blue_rug
red_rug
white_rug
yellow_rug
count
tepih
cilim
krpara
tapet
prostirka
deka/pokrivac
otirac
platno
stazica
carpet
crvenitepih
neznam
tapison
rucnik
tepig
podmetac
tapetic

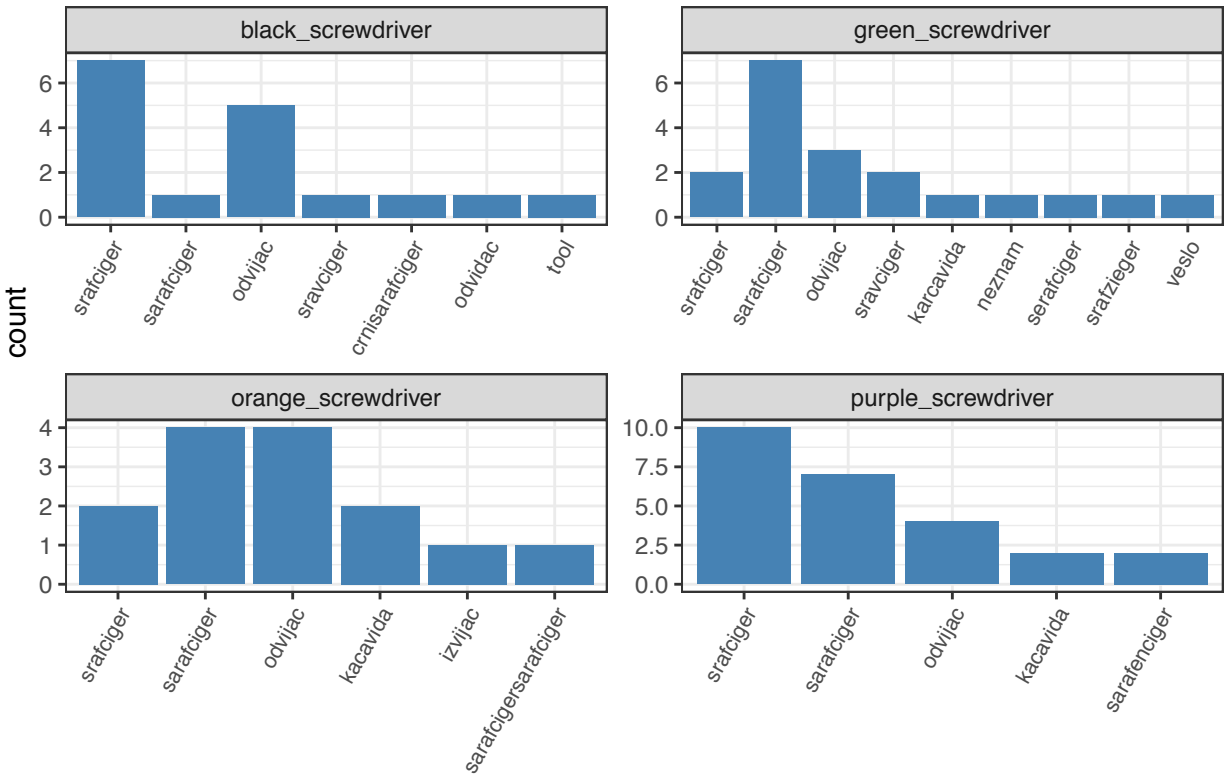
```
##
## [[81]]
```


ruler
blue_ruler
red_ruler
white_ruler
yellow_ruler
count
ravnalo
lenjir
mjerilo
mjrilo
plavoravnalo
ruler
linijar
metar
crta
redalica
ravnala
daska
neznam
vladar

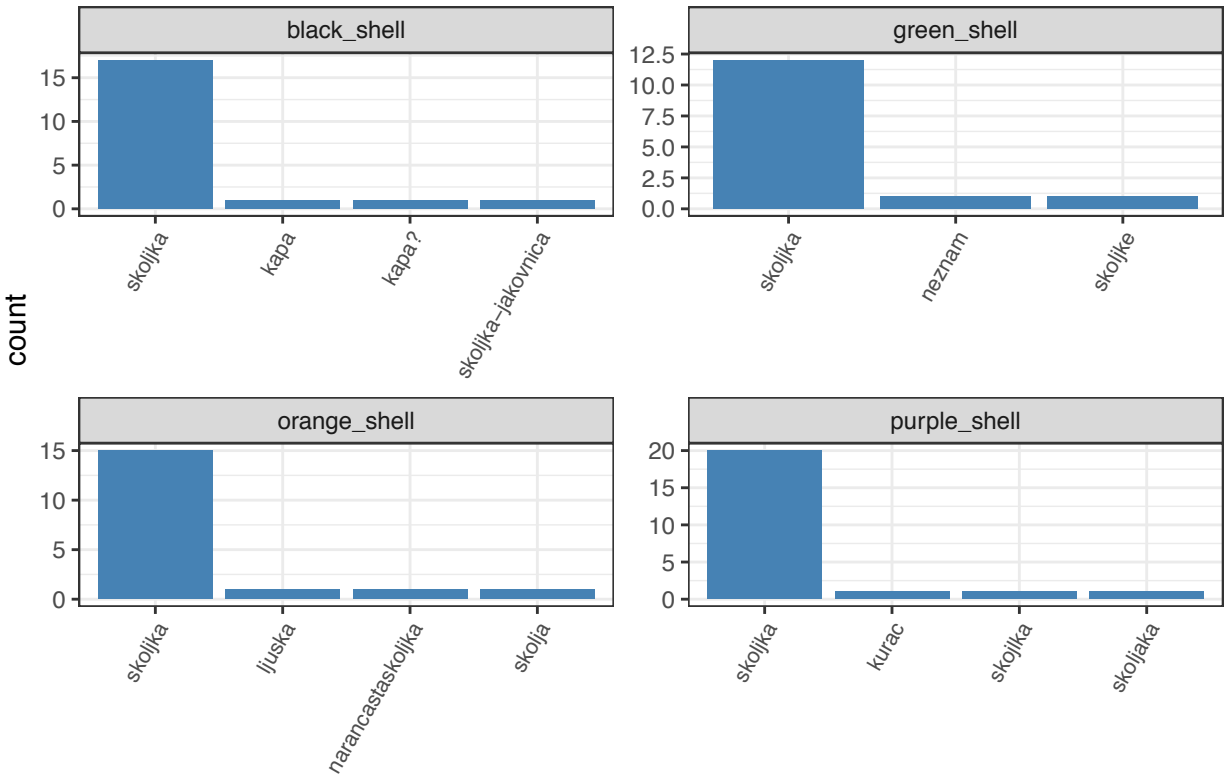
```
##
## [[82]]
```


scarf
blue_scarf
red_scarf
white_scarf
yellow_scarf
count
sal
marama
esarpa
plavisal
sal(3slova)
salokovrata
salzavrat
zaštonemogudapisemreciodtrislovaililkrace?
crvenisal
sal...
salbre
salzavrat
schal
shal
belisal
bijelisal
sal,sal
zimskisal
sal.
salzimski
zutisal
zutsal

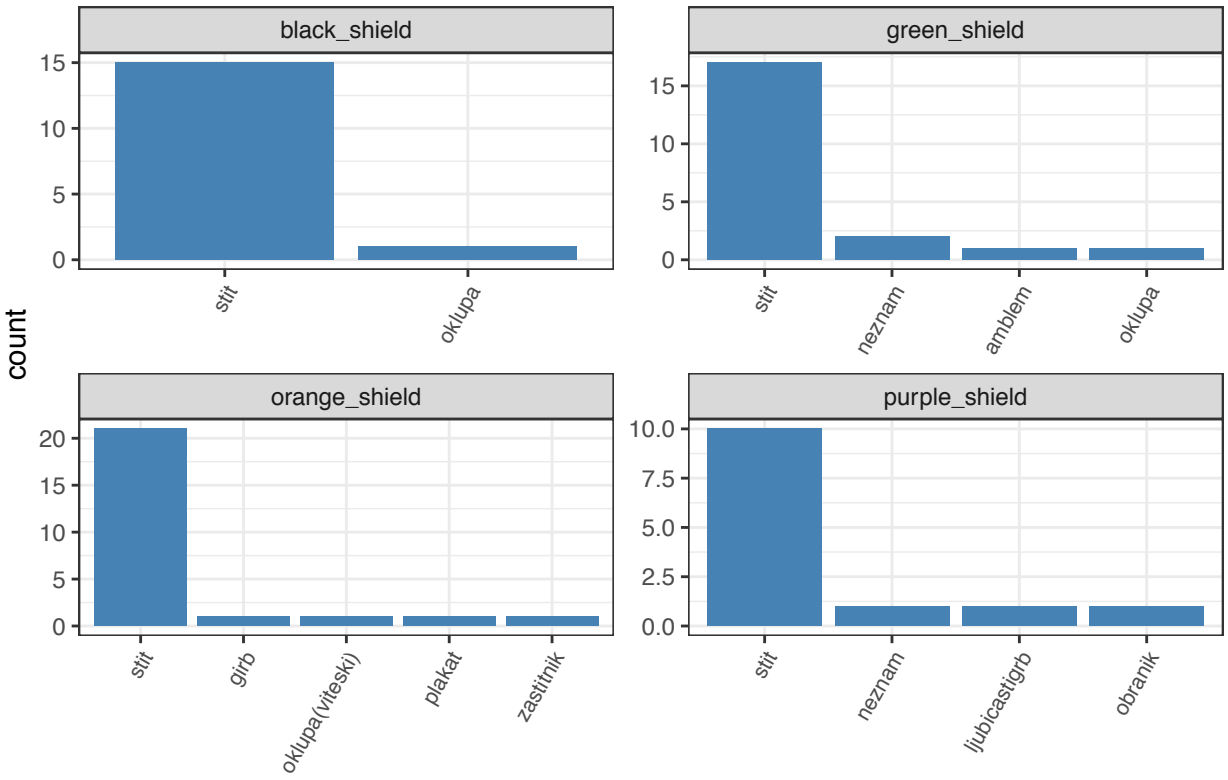
```
##
## [[83]]
```


screwdriver
black_screwdriver
green_screwdriver
orange_screwdriver
purple_screwdriver
count
srafciger
sarafciger
odvijac
sravciger
crnisarafciger
odvidac
tool
karcavida
neznam
serafciger
srafzieger
veslo
kacavida
izvijac
sarafcigersarafciger
sarafenciger

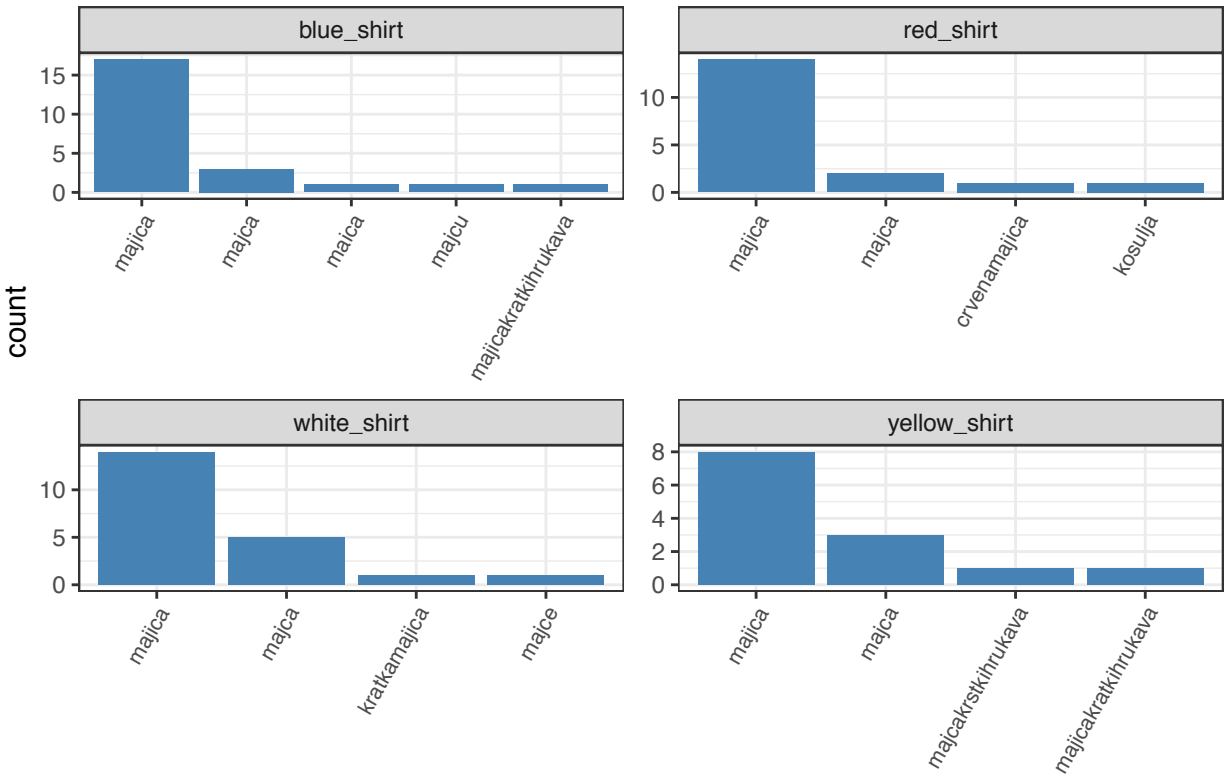
```
##
## [[84]]
```


shell
black_shell
green_shell
orange_shell
purple_shell
count
skoljka
kapa
kapa?
skoljka-jakovnica
skoljka
neznam
skoljke
skoljka
ljuska
narancastaskoljka
skolja
skoljka
kurac
skoljka
skoljaka

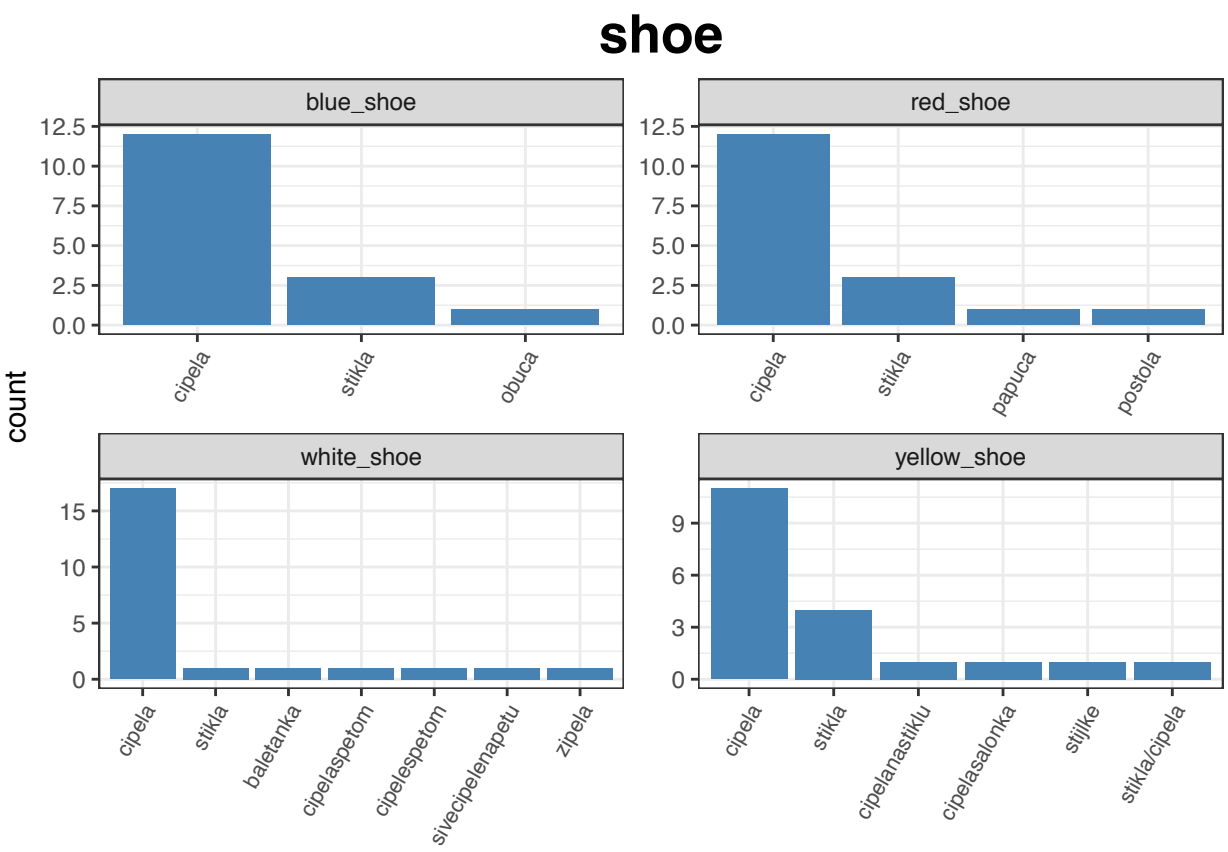
```
##
## [[85]]
```


shield
black_shield
green_shield
orange_shield
purple_shield
count
stit
oklupa
neznam
amblem
girb
oklupa(viteski)
plakat
zastitnik
ljubicastigrb
obranik

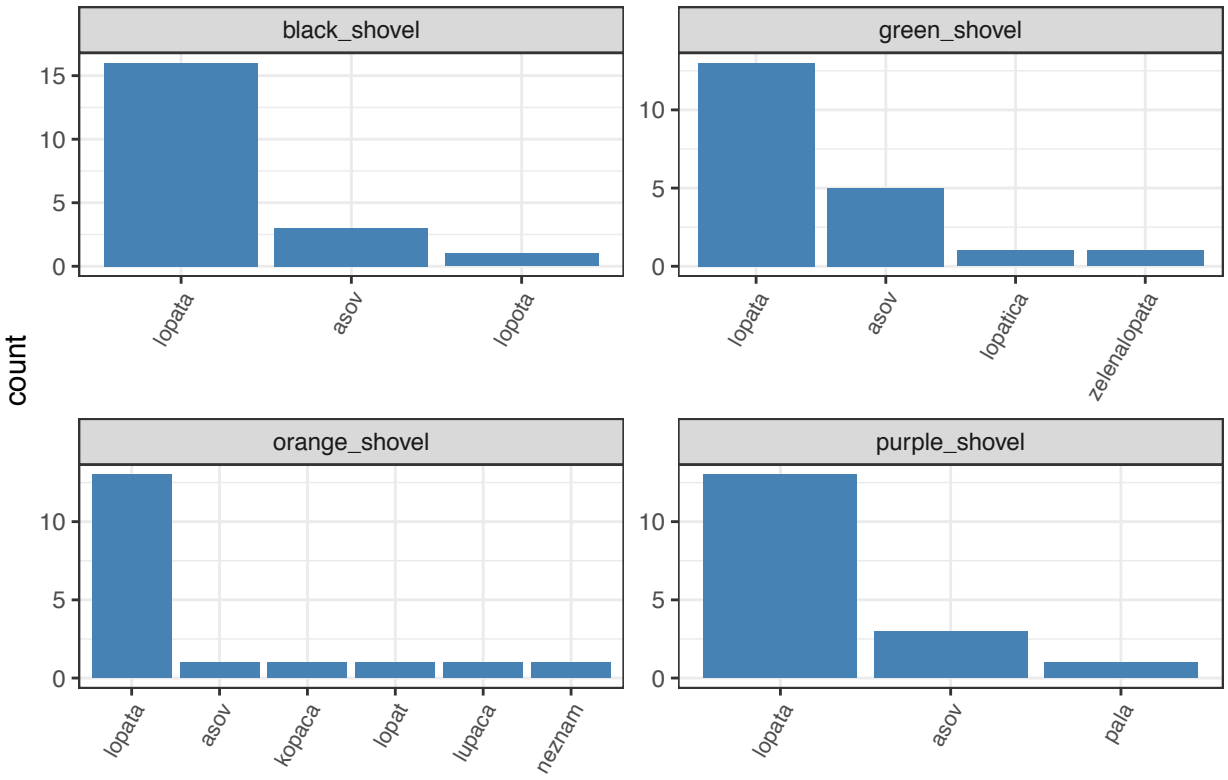
```
##
## [[86]]
```


shirt
blue_shirt
red_shirt
white_shirt
yellow_shirt
count
majica
majca
maica
majcu
majicakratkihrukava
majica
majca
crvenamajica
kosulja
majica
majca
kratkamajica
majce
majica
majca
majicakrstkihrukava
majicakratkihrukava

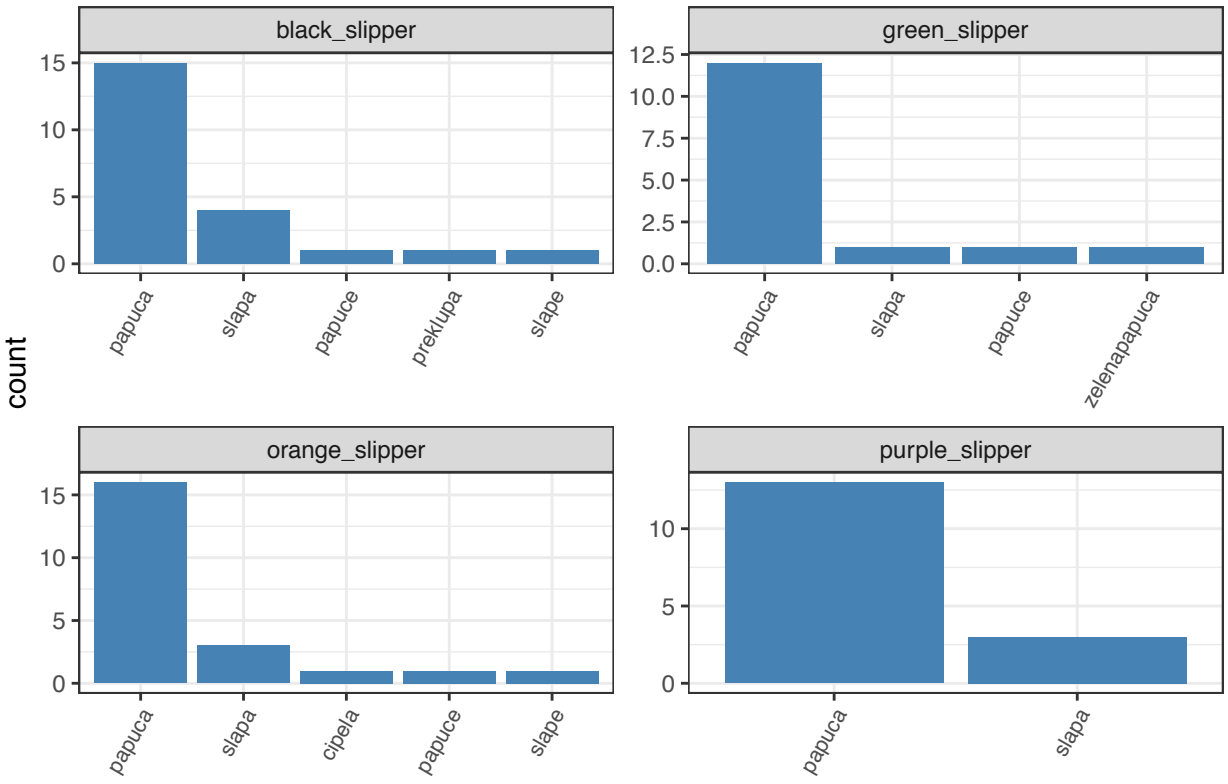
```
##
## [[87]]
```


shoe
blue_shoe
red_shoe
white_shoe
yellow_shoe
count
cipela
stikla
obuca
papuca
postola
baletanka
cipelaspetom
cipelespetom
sivecipelenapetu
zipela
cipelanastiklu
cipelasalonka
stijlke
stikla/cipela

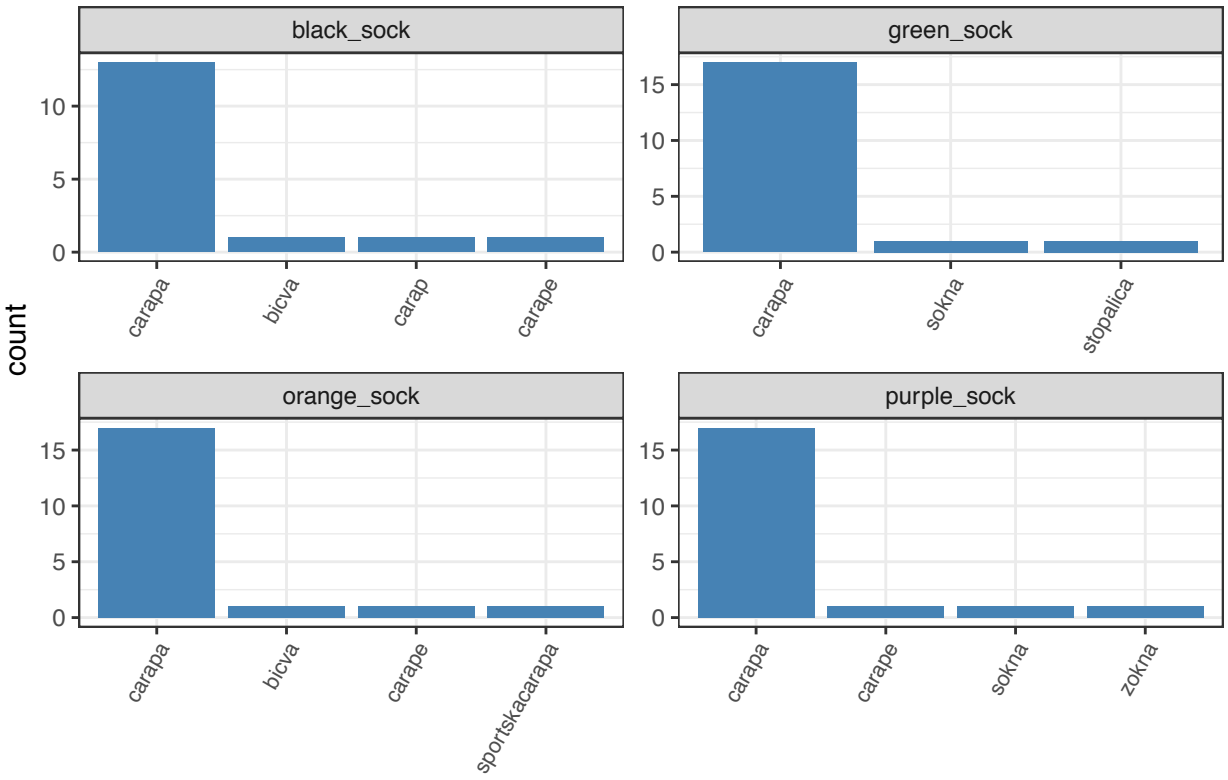
```
##
## [[88]]
```


shovel
black_shovel
green_shovel
orange_shovel
purple_shovel
count
0
5
10
15
lopata
asov
lopota
lopatica
zelenalopata
kopaca
lopat
lupaca
neznam
pala

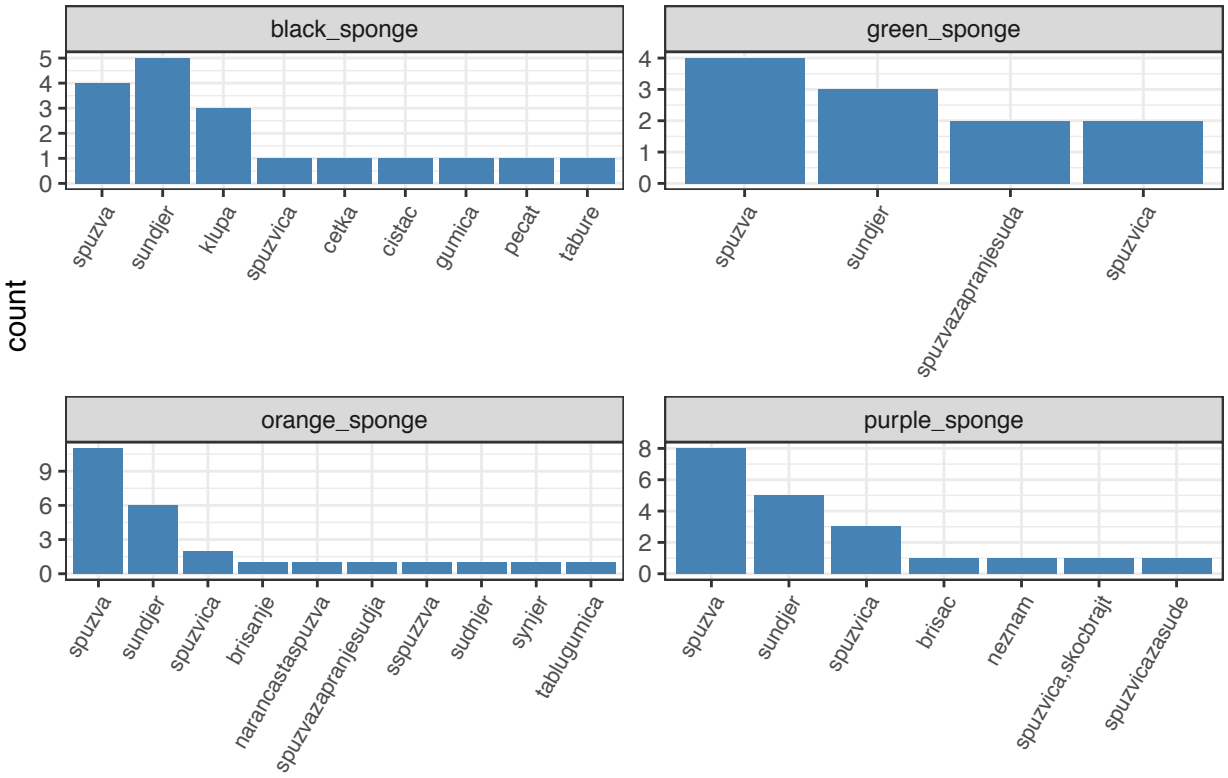
```
##
## [[89]]
```


slipper
black_slipper
green_slipper
orange_slipper
purple_slipper
count
papuca
slapa
papuce
preklupa
slape
zelenapapuca
cipela

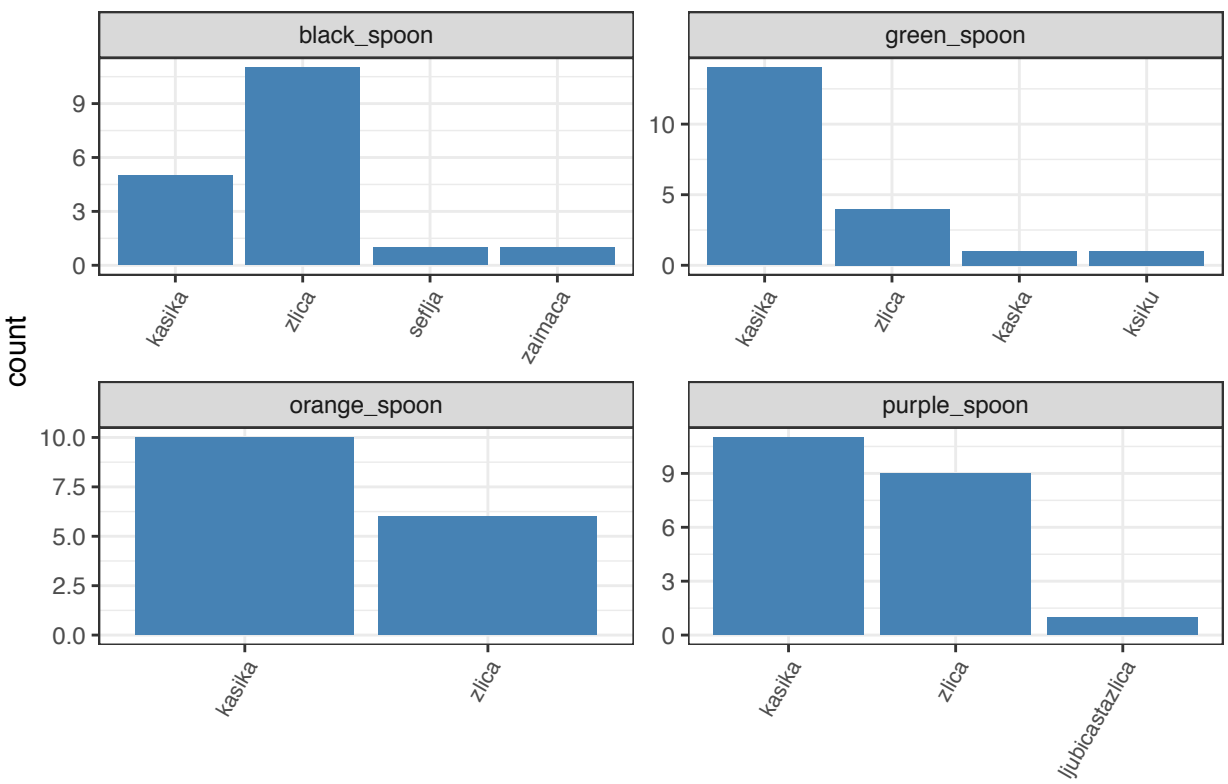
```
##
## [[90]]
```


sock
black_sock
green_sock
orange_sock
purple_sock
count
0
5
10
15
carapa
bicva
carap
carape
sokna
stopalica
sportskacarapa
zokna

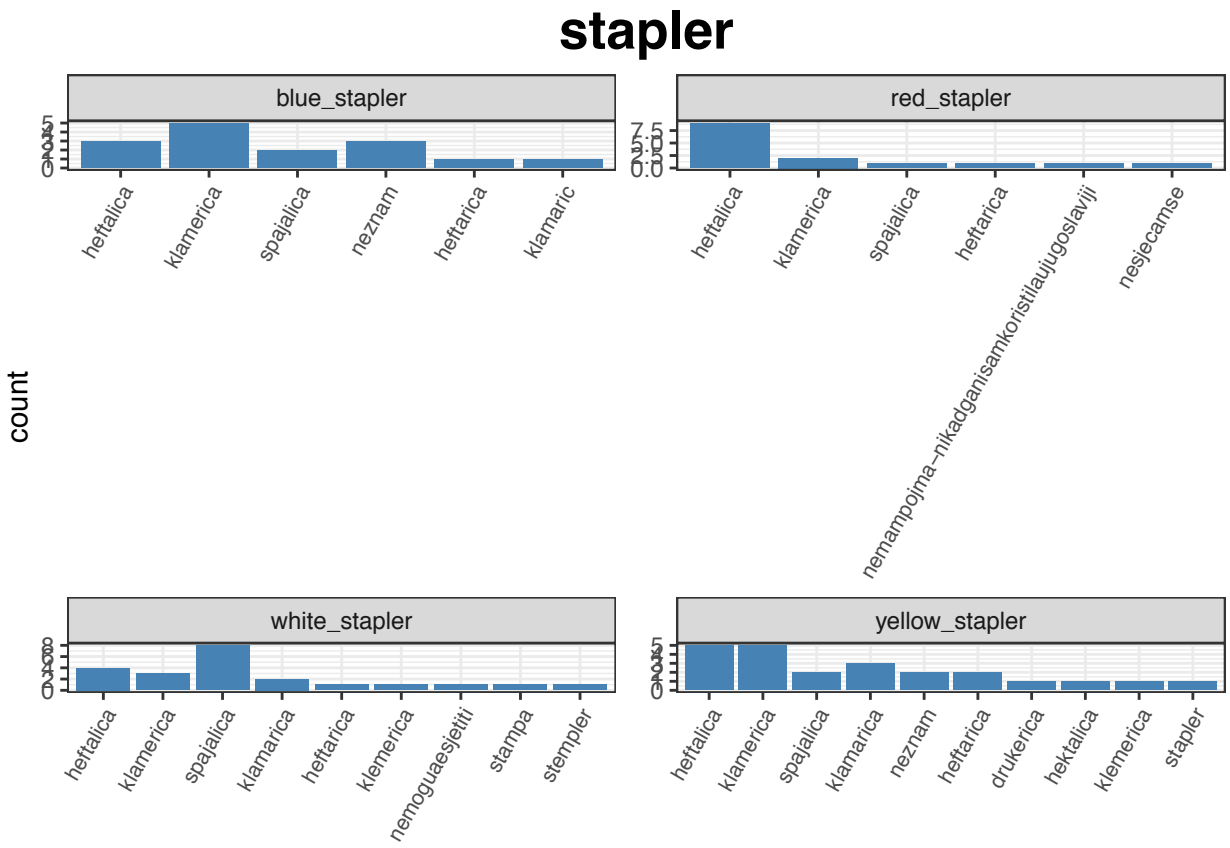
```
##
## [[91]]
```


sponge
black_sponge
green_sponge
orange_sponge
purple_sponge
count
spuzva
sundjer
klupa
spuzvica
cetka
cistac
gumica
pecat
tabure
spuzvazapranjesuda
brisanje
narancastaspuzva
spuzvazapranjesudja
sspuzzva
sudnjer
synjer
tablugumica
brisac
neznam
spuzvica.skocbrajt
spuzvicazasude

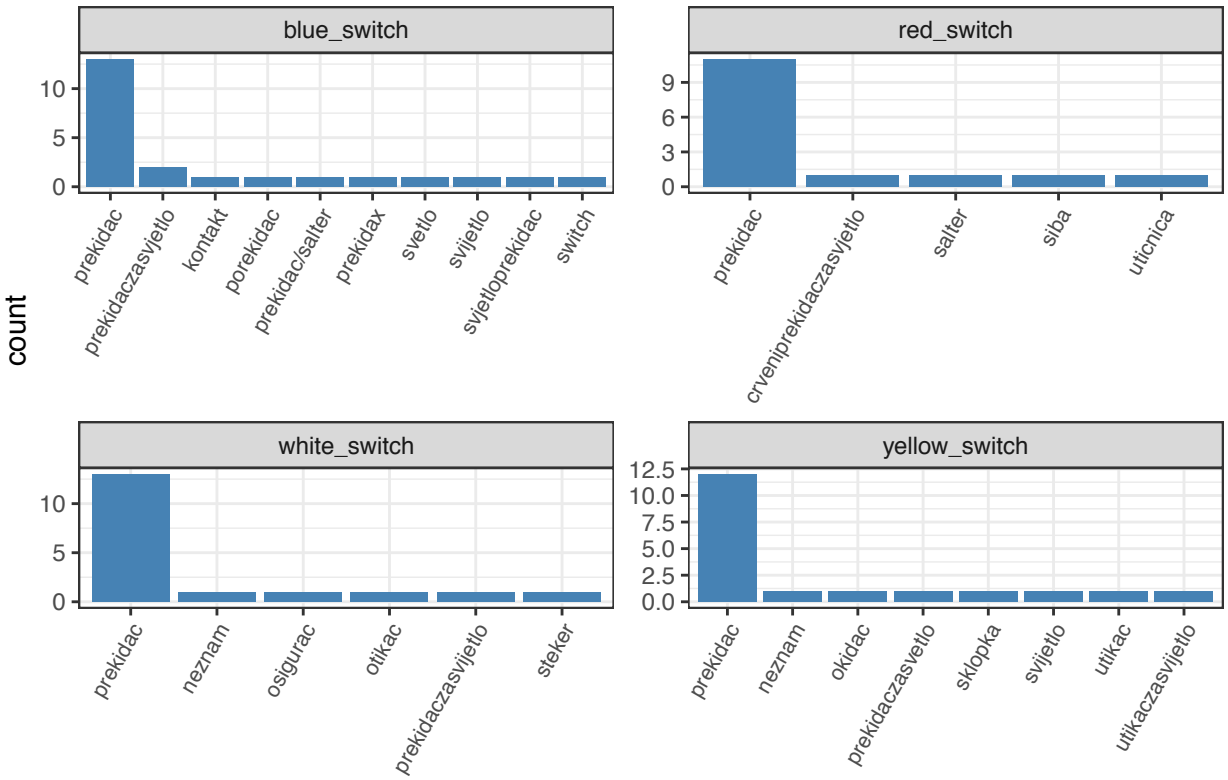
```
##
## [[92]]
```


spoon
black_spoon
green_spoon
orange_spoon
purple_spoon
count
kasika
zlica
seflja
zalimaca
kaska
ksiku
ljubicastazlica

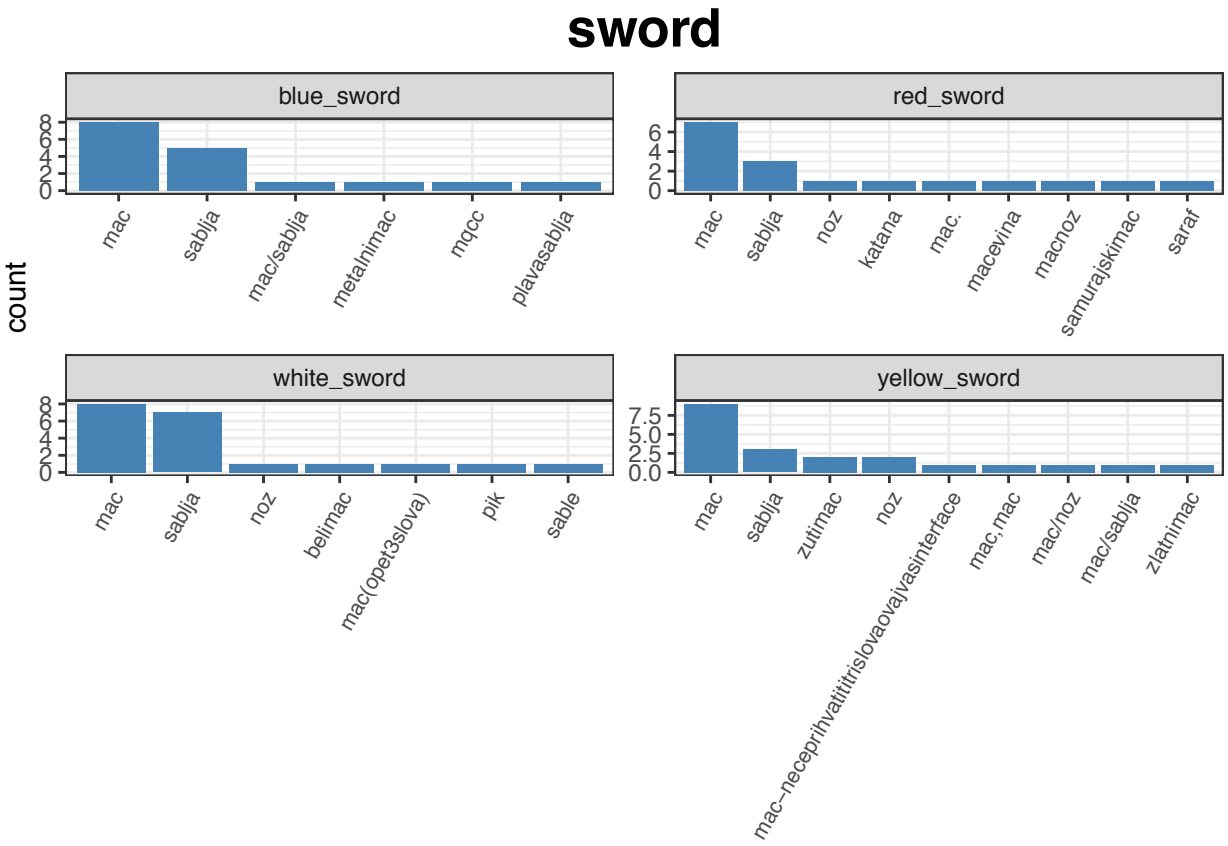
```
##
## [[93]]
```


stapler
blue_stapler
red_stapler
white_stapler
yellow_stapler
count
heftalica
klamerica
spajalica
neznam
heftarica
klamaric
nemampojma–nikadganisamkoristilaujugoslaviji
nesjecamse
klamarica
nemoguaesjetiti
stampa
stempler
drukerica
hektalica
klemerica
stapler

```
##
## [[94]]
```


switch
blue_switch
red_switch
white_switch
yellow_switch
count
prekidac
prekidaczasvijetlo
kontakt
porekidac
prekidac/salter
prekidax
svetlo
svijetlo
svjetloprekidac
switch
crvenoprekidaczasvijetlo
salter
siba
uticnica
neznam
osigurac
otikac
prekidaczasvijetlo
steker
okidac
prekidaczasvetlo
sklopka
svijetlo
utikac
utikaczasvijetlo

```
##
## [[95]]
```


sword
blue_sword
red_sword
white_sword
yellow_sword
count
mac
sablja
mac/sablja
metalnimac
mqcc
plavasablja
noz
katana
mac.
macevina
macnoz
samurajskimac
saraf
belimac
mac(opet3slova)
plk
sable
zutimac
mac-neceprihvatititrislovaovajvasinterface
mac,mac
mac/noz
mac/sablja
zlatnimac

```
##
## [[96]]
```

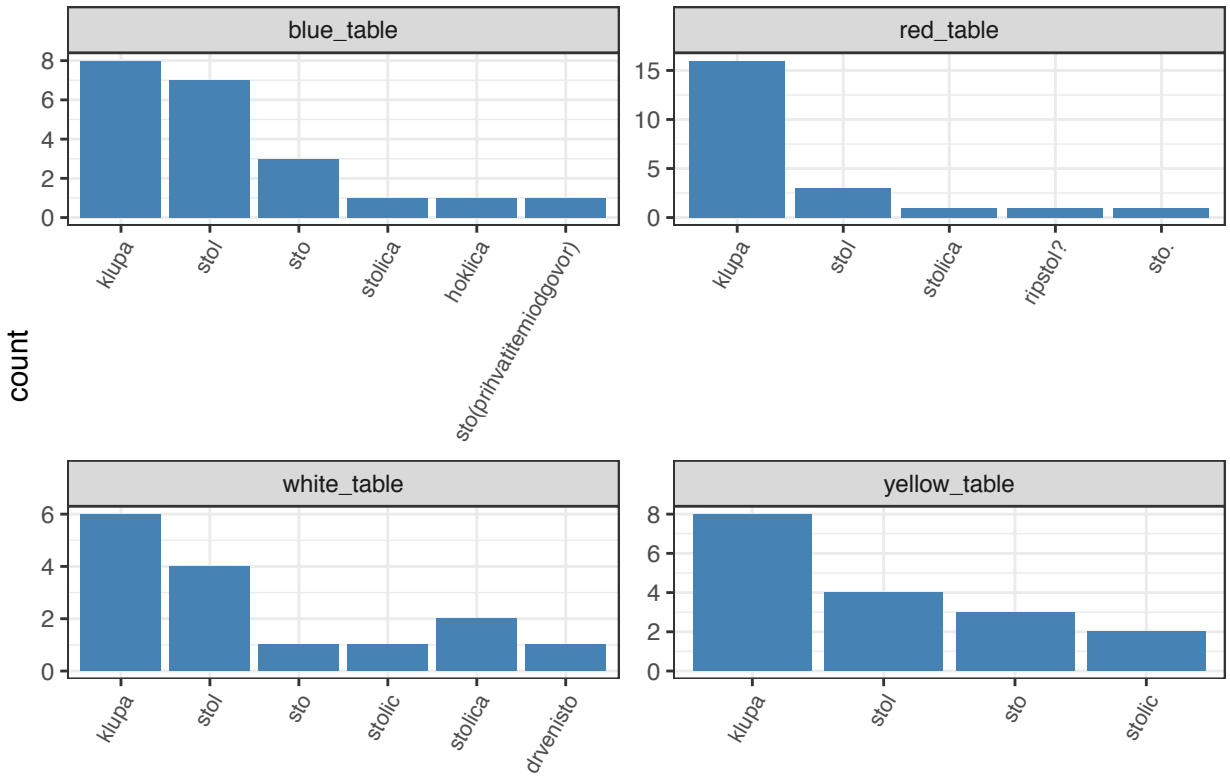
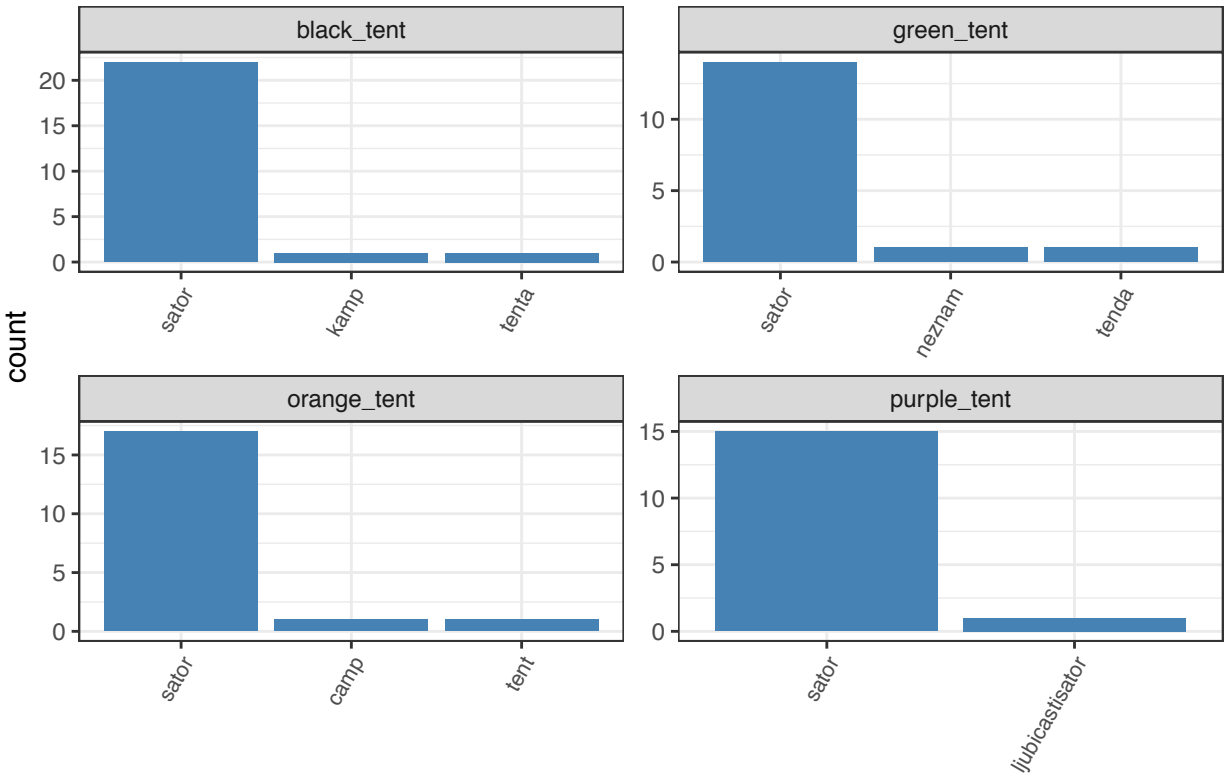
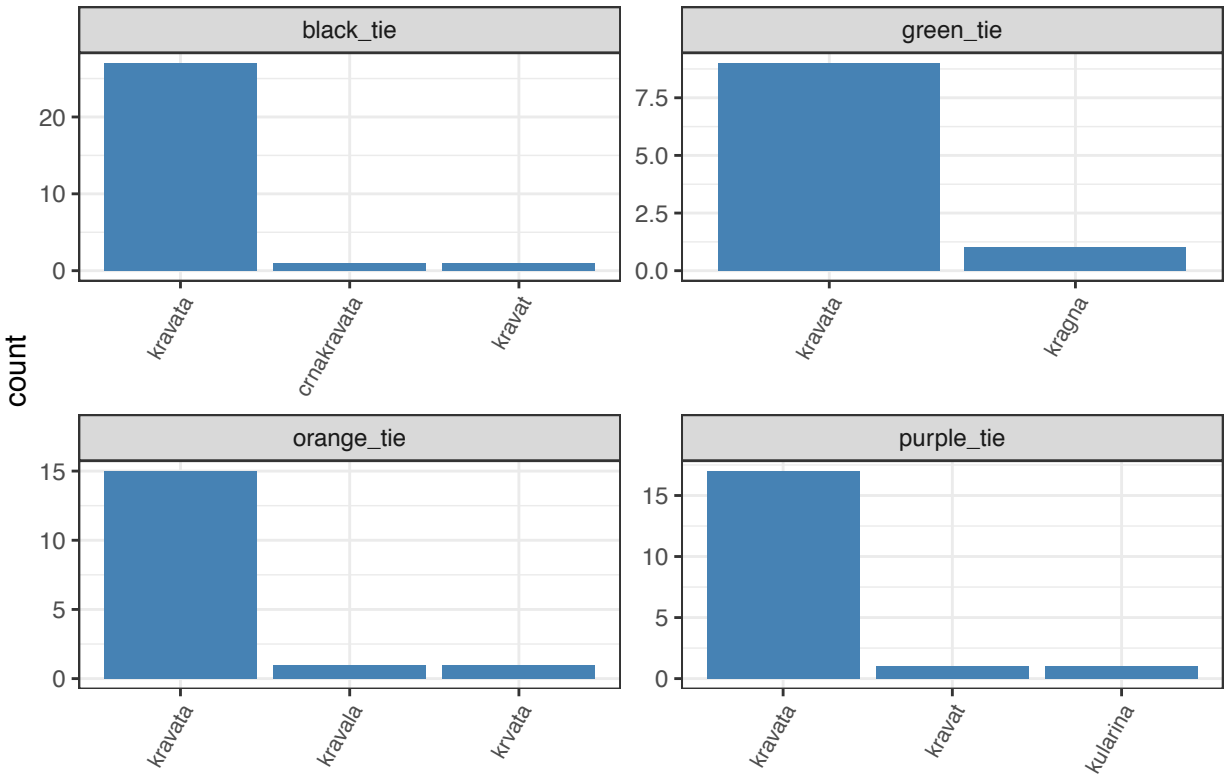


table
blue_table
red_table
white_table
yellow_table
count
klupa
stol
sto
stolica
hoklica
sto(prihvatitemiodgovor)
ripstol?
sto.
stolic
drvenisto

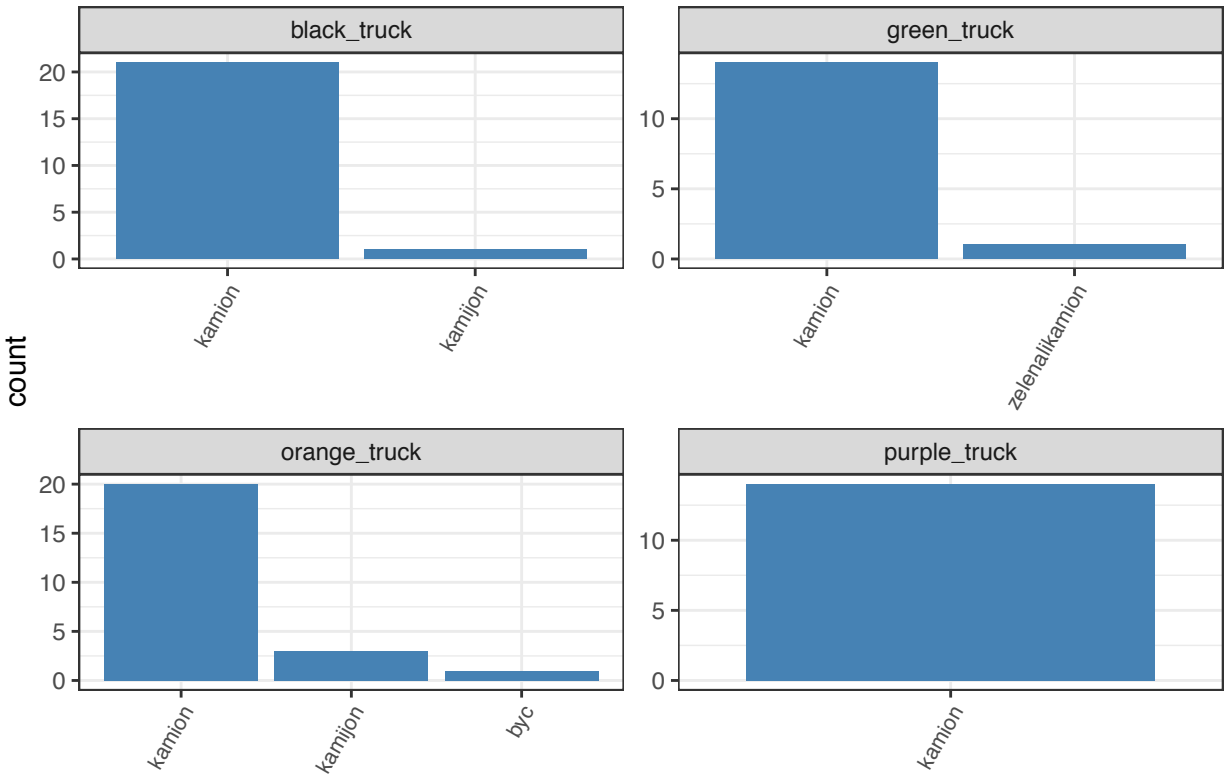
```
##
## [[97]]
```


tent
black_tent
green_tent
orange_tent
purple_tent
count
sator
kamp
tenta
neznam
tenda
camp
tent
ljubicastisator

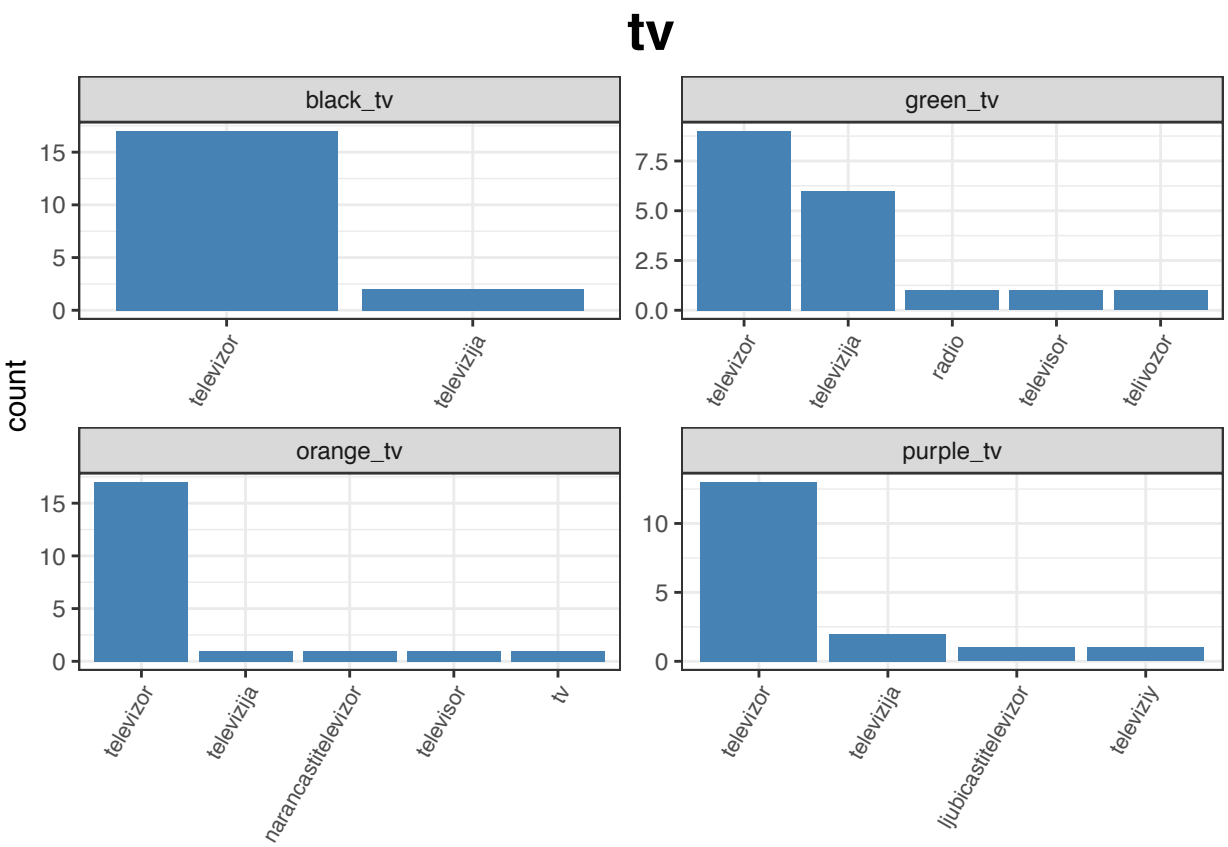
```
##
## [[98]]
```


tie
black_tie
green_tie
orange_tie
purple_tie
count
kravata
crnakravata
kravat
kragna
kravala
krvata
kulairina

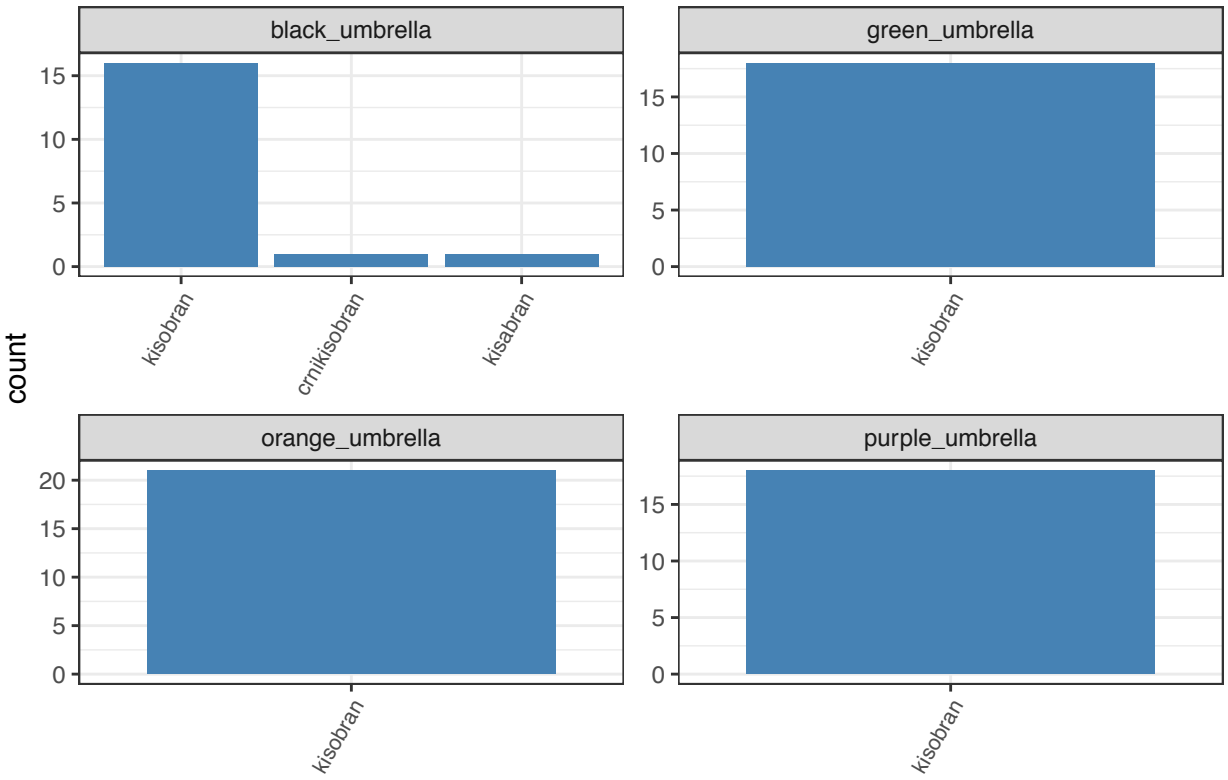
```
##
## [[99]]
```


truck
black_truck
green_truck
orange_truck
purple_truck
count
0
5
10
15
20
kamion
kamijon
zelenalikamion
byc

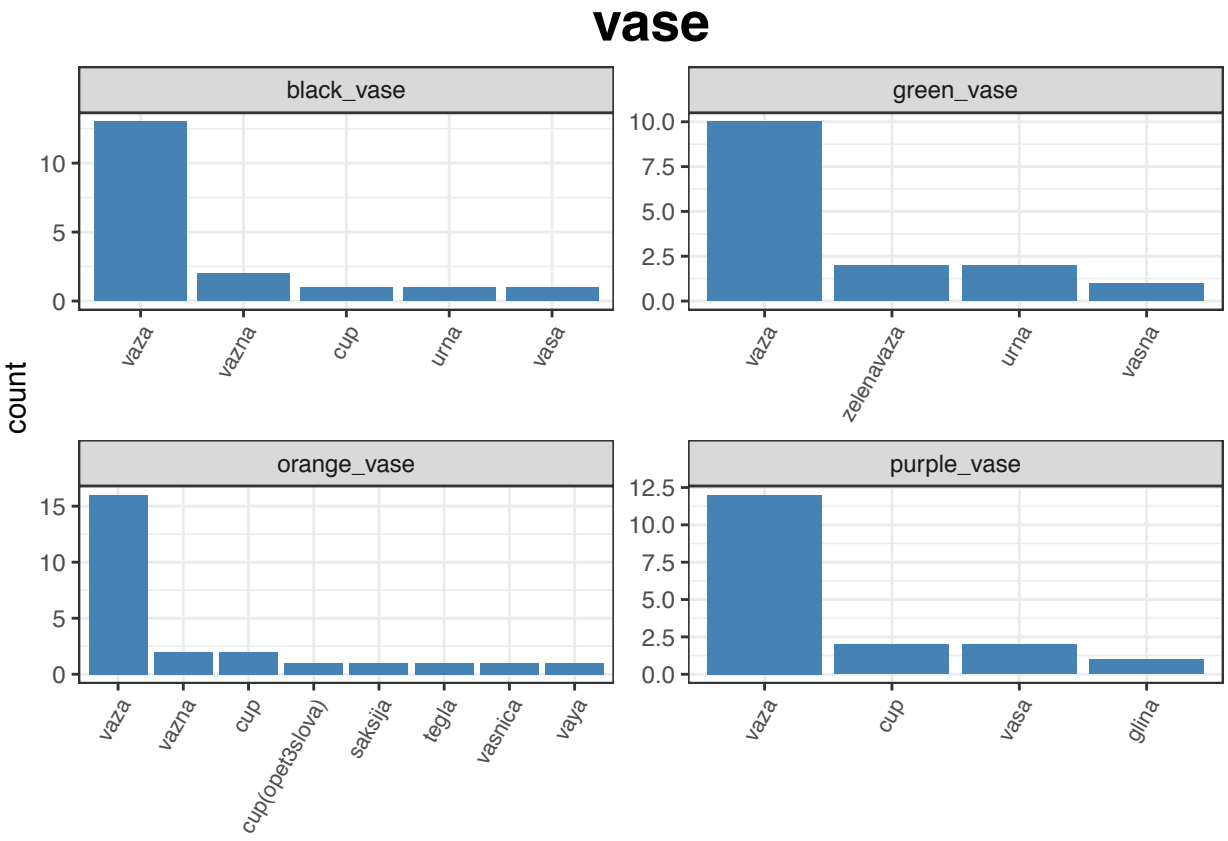
```
##
## [[100]]
```


tv
black_tv
green_tv
orange_tv
purple_tv
count
televizor
televizija
radio
televisor
tellivozor
narancastitelevizor
tv
ljubicastitelevizor
televiziy

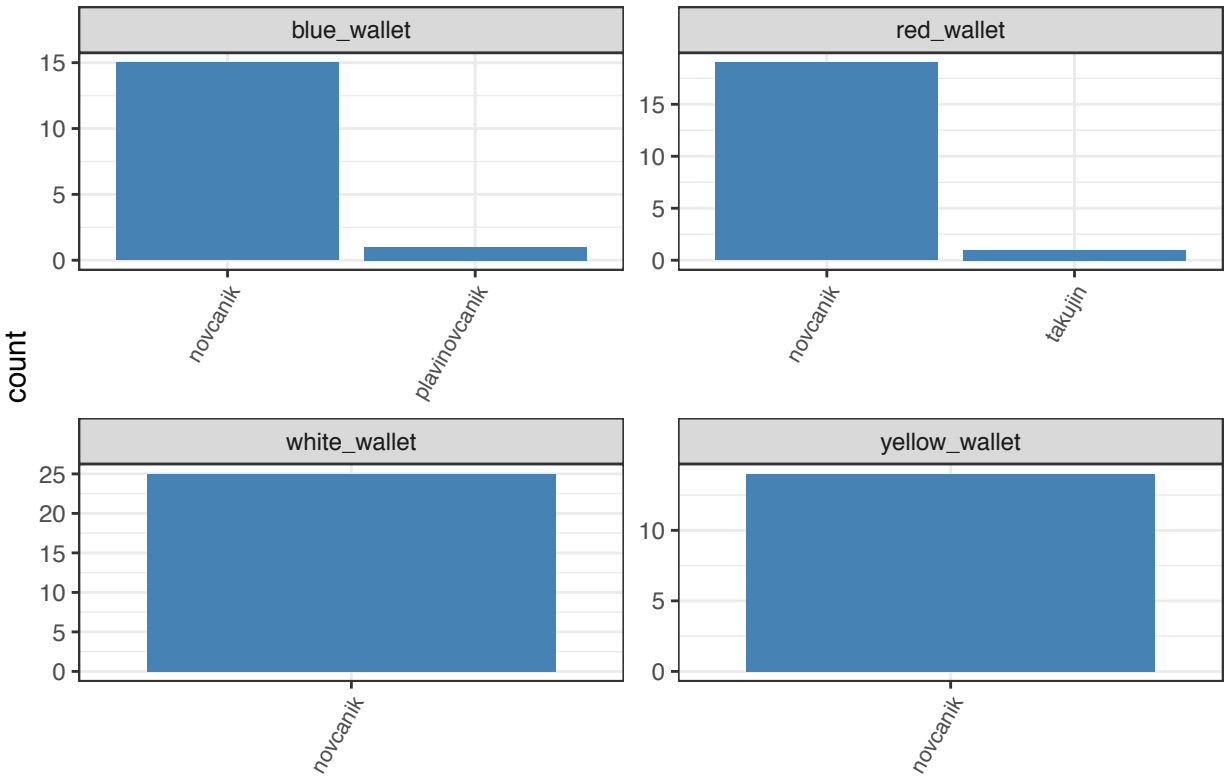
```
##
## [[101]]
```


umbrella
black_umbrella
green_umbrella
orange_umbrella
purple_umbrella
count
kisobran
crnikisobran
kisabran

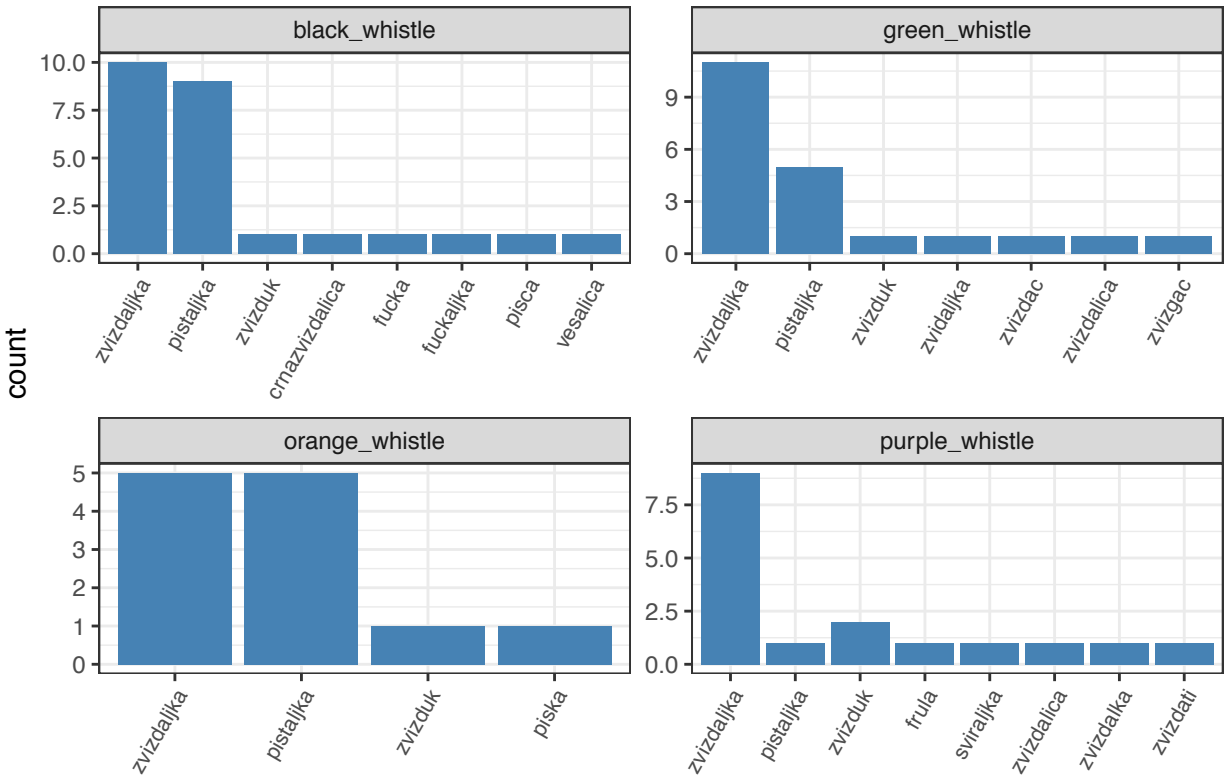
```
##
## [[102]]
```


vase
black_vase
green_vase
orange_vase
purple_vase
count
vaza
vazna
cup
urna
vasa
zelenavaza
vasna
cup(opet3slova)
saksija
tegla
vasnica
vaya
glina

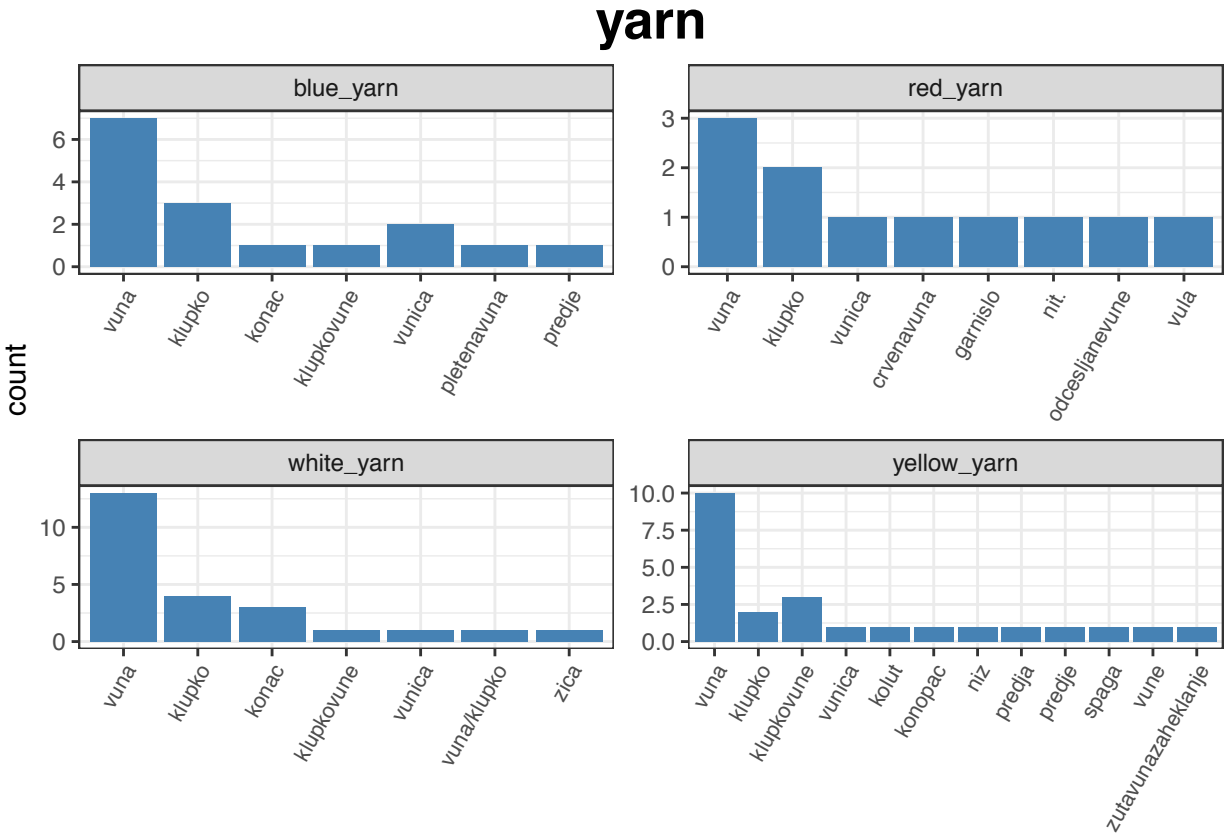
```
##
## [[103]]
```


wallet
blue_wallet
red_wallet
white_wallet
yellow_wallet
count
novcanik
plavinovcanik
takujin

```
##
## [[104]]
```


whistle
black_whistle
green_whistle
orange_whistle
purple_whistle
count
zvizdaljka
pistaljka
zvizduk
crnazvizdalica
fucka
fuckaljka
pisca
vesalica
zvizdaljka
pistaljka
zvizduk
zvidaljka
zvizdac
zvizdalica
zvizgac
zvizdaljka
pistaljka
zvizduk
piska
zvizdaljka
pistaljka
zvizduk
frula
sviraljka
zvizdalica
zvizdalka
zvizdati

```
##
## [[105]]
```


yarn
count
blue_yarn
red_yarn
white_yarn
yellow_yarn
vuna
klupko
konac
klupkovune
vunica
pletenavuna
predje
crvenavuna
garnislo
nit.
odcesljanevune
vuna/klupko
zica
kolut
konopac
niz
predja
spaga
vune
zutavunazaheklanje