

Kickstarter_Casual_Analysis

Harry Emeric

03/12/2019

```
#data1 <- read.csv("../data/cleaned_kickstarterFinal(1).csv", sep=",")
```

```
data_clean <- read.csv("../data/kickstarter_data.csv", sep=",")
```

```
#EDA
```

```
head(data_clean)
```

```
##   avg_reward_description blurb_length blurb_word_count category_art
## 1                      95           66              13           0
## 2                     225          121              19           0
## 3                     123           57              11           0
## 4                     193          132              22           0
## 5                      64          118              18           0
## 6                     179          106              14           0
##   category_comics category_crafts category_dance category_design
## 1                0              0              0              0
## 2                0              0              0              0
## 3                0              0              0              0
## 4                0              0              0              0
## 5                0              0              0              0
## 6                0              0              0              0
##   category_fashion category_film_video category_food category_games
## 1                0                  0              0              0
## 2                0                  0              0              0
## 3                0                  0              0              0
## 4                0                  0              1              0
## 5                1                  0              0              0
## 6                1                  0              0              0
##   category_journalism category_music category_photography
## 1                   0              1              0
## 2                   0              0              1
## 3                   0              1              0
## 4                   0              0              0
## 5                   0              0              0
## 6                   0              0              0
##   category_publishing category_technology category_theater country_at
## 1                   0                  0              0              0
## 2                   0                  0              0              0
## 3                   0                  0              0              0
## 4                   0                  0              0              0
## 5                   0                  0              0              0
## 6                   0                  0              0              0
##   country_au country_be country_ca country_ch country_de country_dk
## 1           0           0           0           0           0           0
## 2           0           0           0           0           0           0
## 3           0           0           0           0           0           0
## 4           0           0           0           0           0           0
```

## 5	0	0	0	0	0	0
## 6	0	0	0	0	0	0
##	country_es	country_fr	country_gb	country_hk	country_ie	country_it
## 1	0	0	0	0	0	0
## 2	0	0	0	0	0	0
## 3	0	0	0	0	0	0
## 4	0	0	1	0	0	0
## 5	0	0	0	0	0	0
## 6	0	0	0	0	0	0
##	country_jp	country_lu	country_mx	country_nl	country_no	country_nz
## 1	0	0	0	0	0	0
## 2	0	0	0	0	0	0
## 3	0	0	0	0	0	0
## 4	0	0	0	0	0	0
## 5	0	0	0	0	0	0
## 6	0	0	0	0	1	0
##	country_se	country_sg	country_us	creator_project_no		
## 1	0	0	1	1		
## 2	0	0	1	1		
## 3	0	0	1	1		
## 4	0	0	0	1		
## 5	0	0	1	1		
## 6	0	0	0	1		
##	duration_campaign_creation_to_launch					
## 1			0			
## 2			231			
## 3			30			
## 4			2			
## 5			6			
## 6			32			
##	duration_campaign_launch_to_deadline	faq	has_faq	has_one_dollar_reward		
## 1		84	0	0		0
## 2		36	0	0		0
## 3		21	0	0		1
## 4		45	2	1		0
## 5		25	0	0		0
## 6		29	0	0		0
##	is_asking_for_help	is_first_project_from_creator	is_project_we_love			
## 1	0		1			0
## 2	0		1			0
## 3	0		1			0
## 4	0		1			1
## 5	0		1			0
## 6	0		1			0
##	name_length	name_word_count	rewards_levels	rewards_max	rewards_mean	
## 1	35	8	4	100	43	
## 2	41	5	7	500	196	
## 3	30	5	11	450	104	
## 4	32	5	16	1968	349	
## 5	19	2	3	30	18	
## 6	53	11	17	655	152	
##	rewards_min	state	usd_goal	usd_pledged		
## 1	10	0	6000	0		
## 2	10	1	2900	2938		

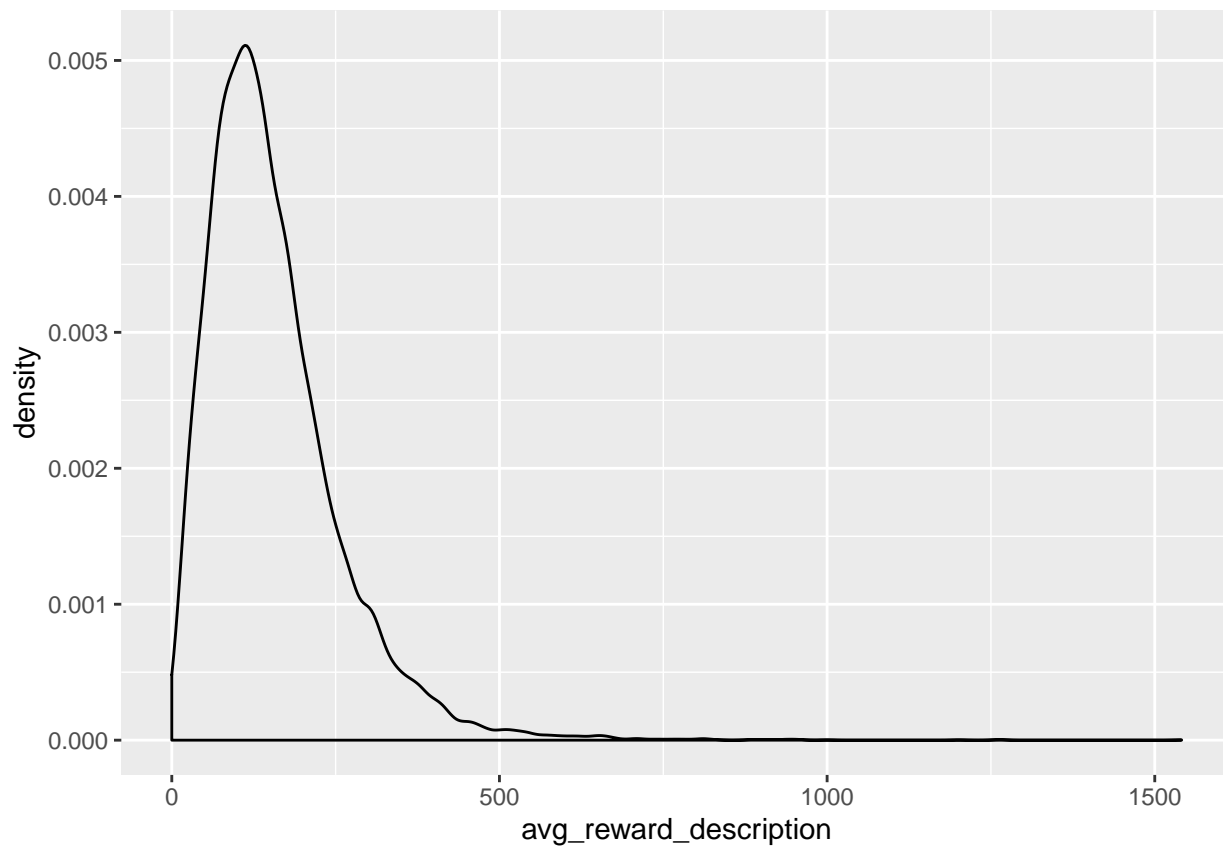
```
## 3      1      0      500      10
## 4     16      1    79131    133970
## 5      5      0     2000      2
## 6     17      1     6248     8706
```

```
N <- dim(data_clean)[1]
```

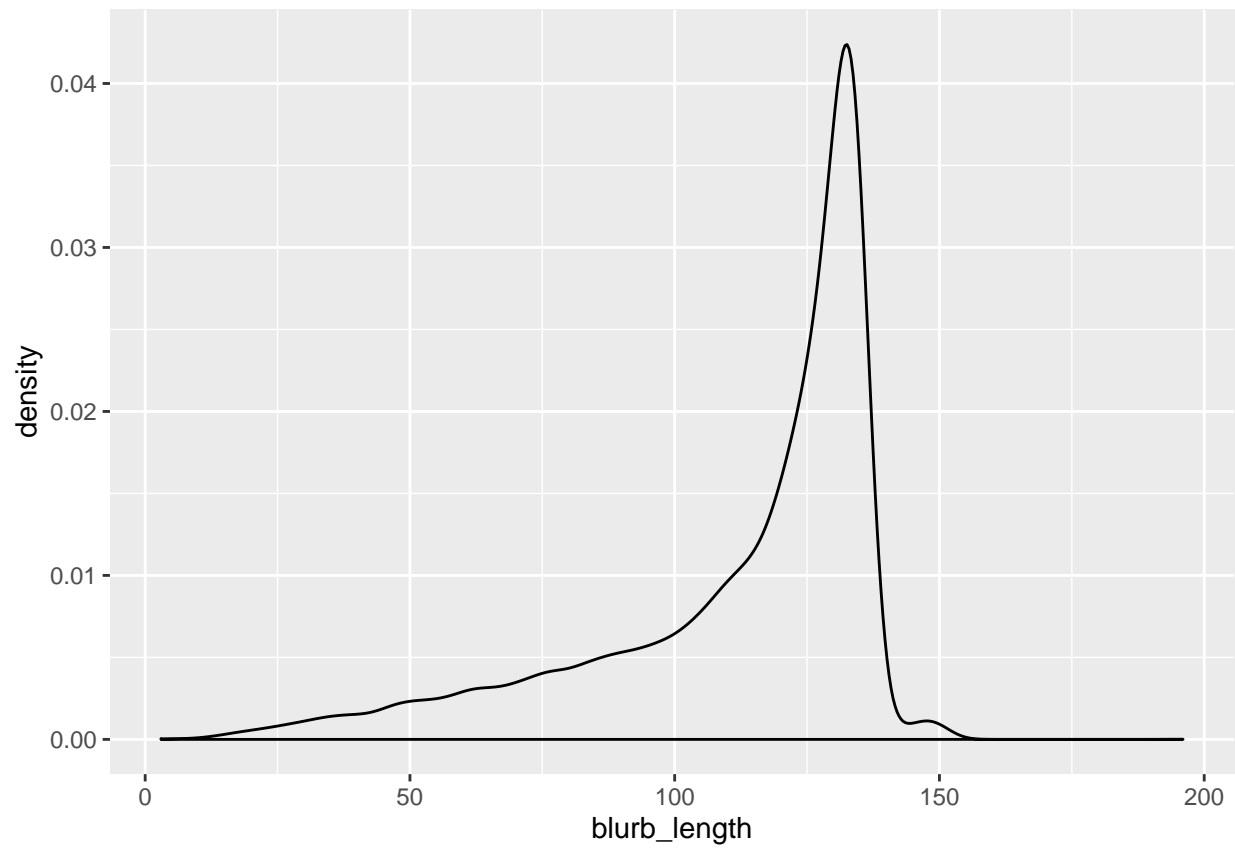
```
data_clean %>% group_by(creator_project_no) %>% summarize(n())
```

```
## # A tibble: 58 x 2
##   creator_project_no `n()`
##   <int> <int>
## 1         1 14652
## 2         2  1647
## 3         3   495
## 4         4   240
## 5         5   127
## 6         6    92
## 7         7    83
## 8         8    47
## 9         9    44
## 10        10    32
## # ... with 48 more rows
```

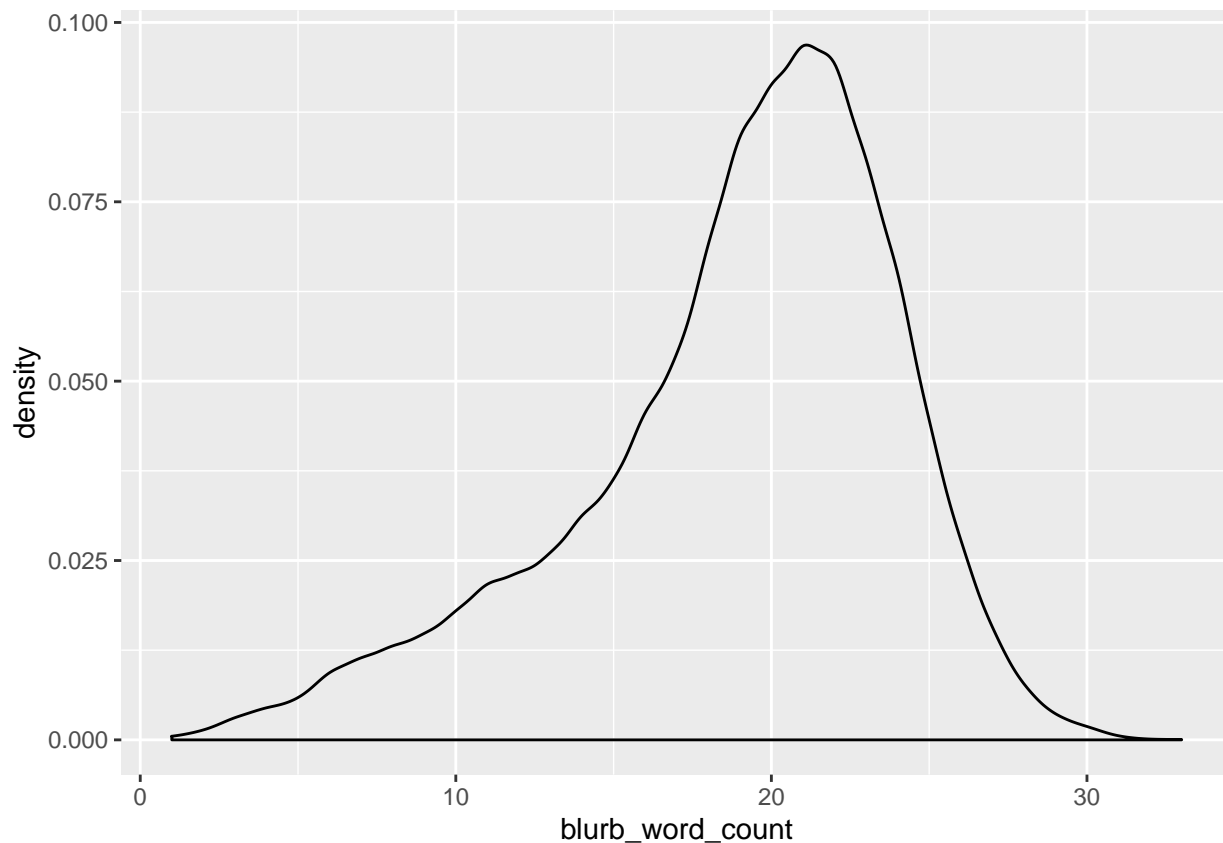
```
ggplot(data=data_clean, aes(x=avg_reward_description)) + geom_density(alpha=0.5)
```



```
ggplot(data=data_clean, aes(x=blurb_length)) + geom_density(alpha=0.5)
```



```
ggplot(data=data_clean, aes(x=blurb_word_count)) + geom_density(alpha=0.5)
```



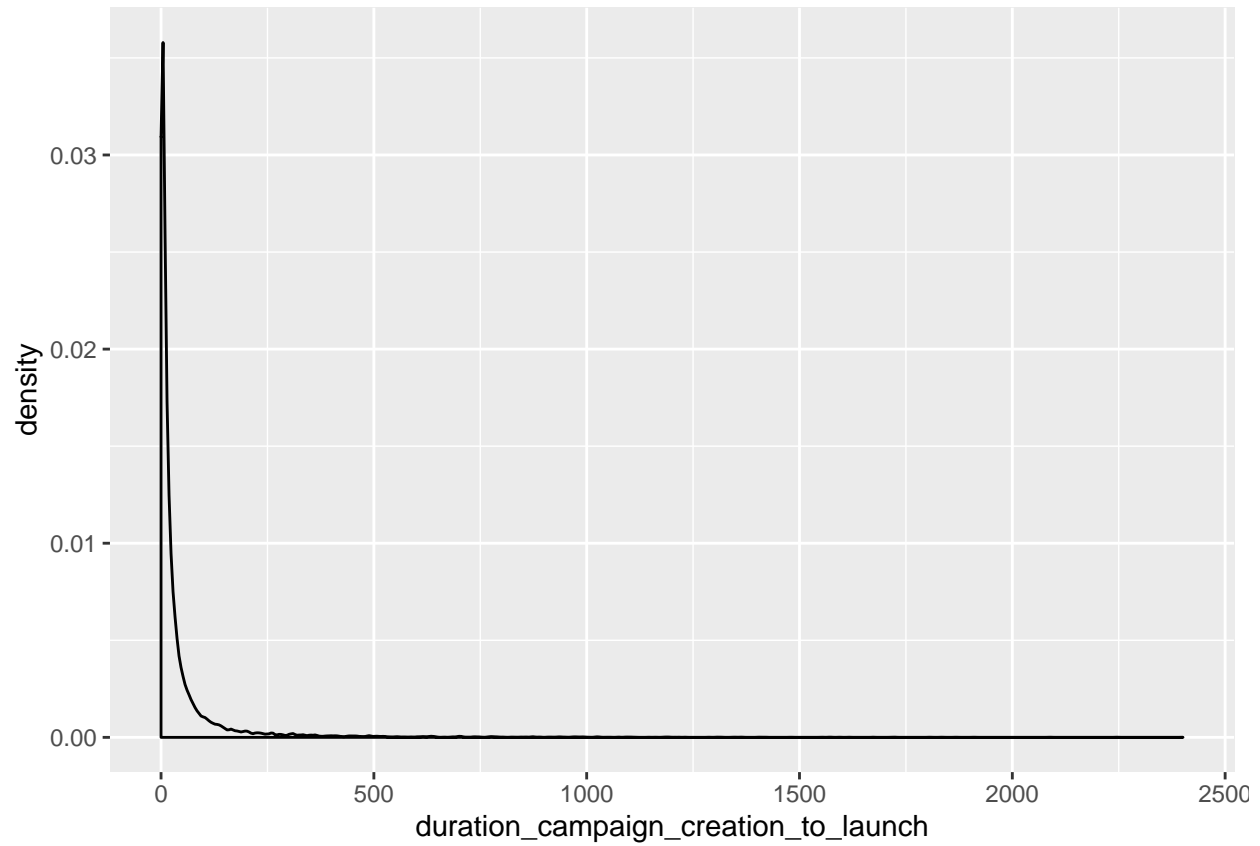
```
categories <- colnames(data_clean)[grepl("^category_", colnames(data_clean))]
data_clean %>% select(categories) %>% colSums()
```

```
##      category_art      category_comics      category_crafts
##      1555            635            380
##      category_dance      category_design      category_fashion
##      199              979            1198
##      category_film_video      category_food      category_games
##      2514            1095            1750
##      category_journalism      category_music      category_photography
##      293              2432            448
##      category_publishing      category_technology      category_theater
##      2059            1630            485
```

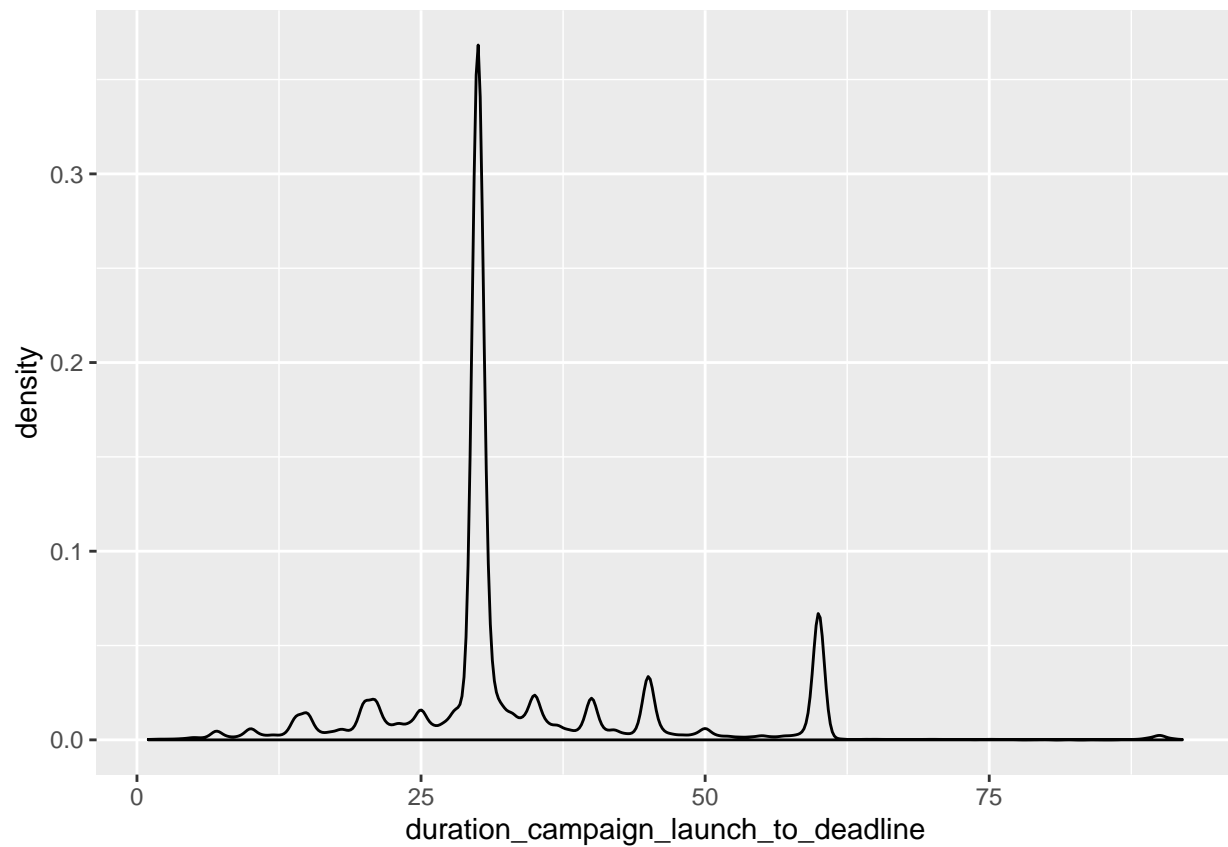
```
countries <- colnames(data_clean)[grepl("^country_", colnames(data_clean))]
data_clean %>% select(countries) %>% colSums()
```

```
## country_at country_au country_be country_ca country_ch country_de
##      36      377      31      707      42      241
## country_dk country_es country_fr country_gb country_hk country_ie
##      56      134      183      1769      77      42
## country_it country_jp country_lu country_mx country_nl country_no
##      170      28      3      139      139      43
## country_nz country_se country_sg country_us
##      78      98      46      13213
```

```
ggplot(data=data_clean, aes(x=duration_campaign_creation_to_launch)) + geom_density(alpha=0.5) #Outlier
```

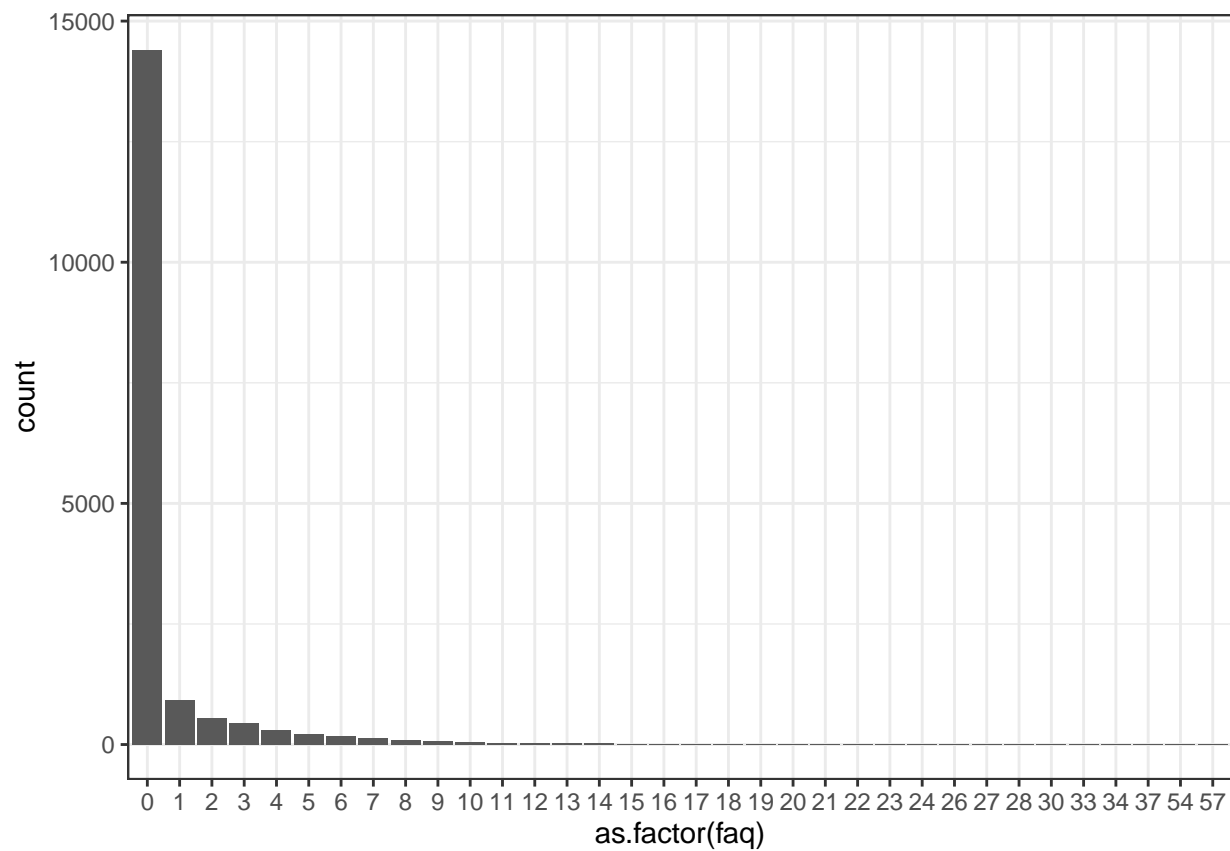


```
ggplot(data=data_clean, aes(x=duration_campaign_launch_to_deadline)) + geom_density(alpha=0.5)
```



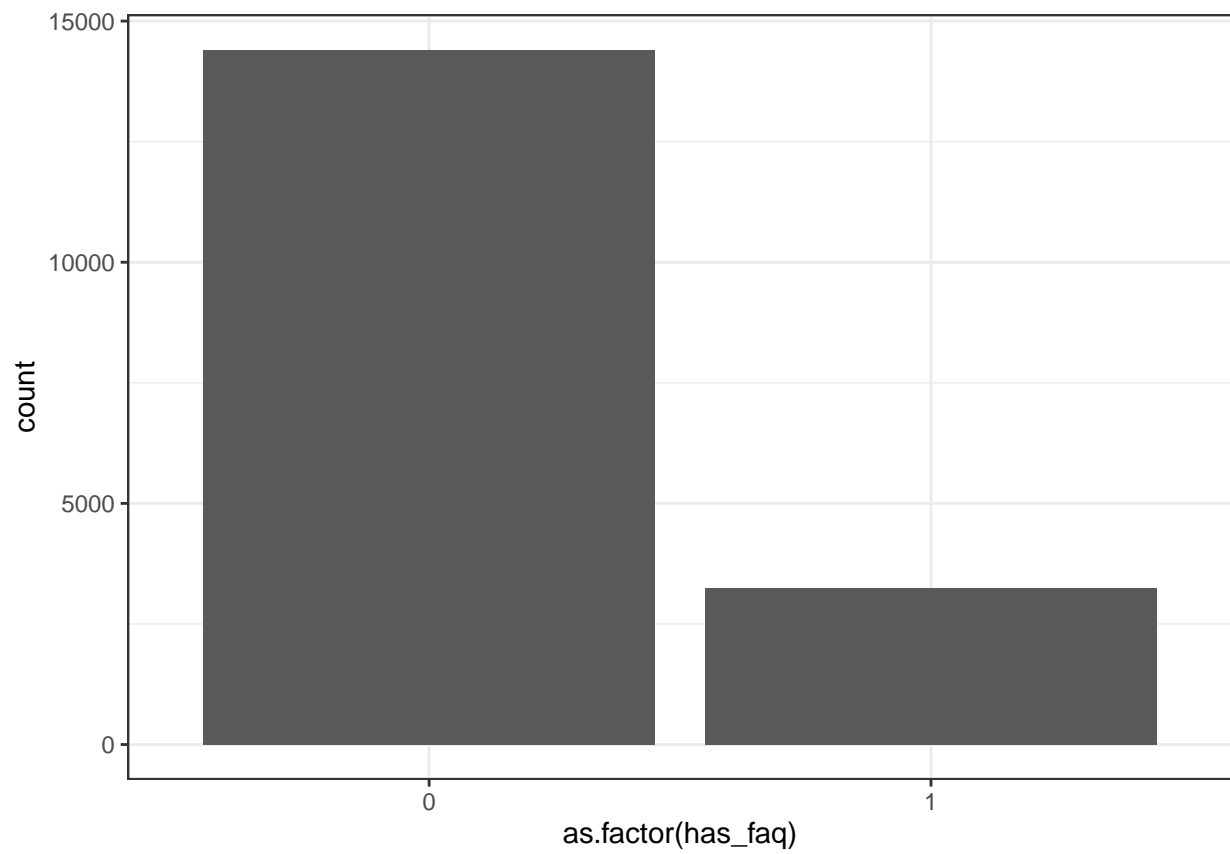
```
ggplot(data=data_clean, aes(x=as.factor(faq))) + geom_histogram(stat="count") + theme_bw()
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



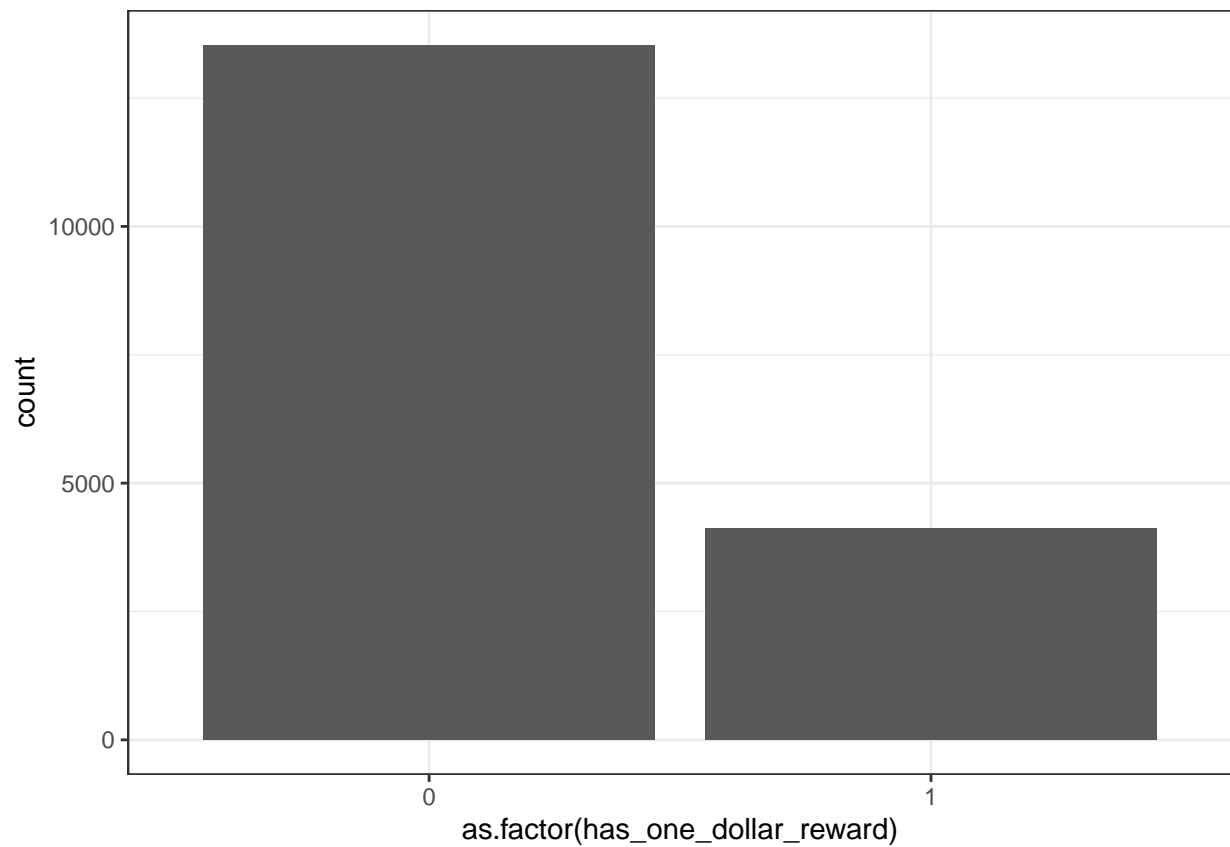
```
ggplot(data=data_clean, aes(x=as.factor(has_faq))) + geom_histogram(stat="count") + theme_bw()
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

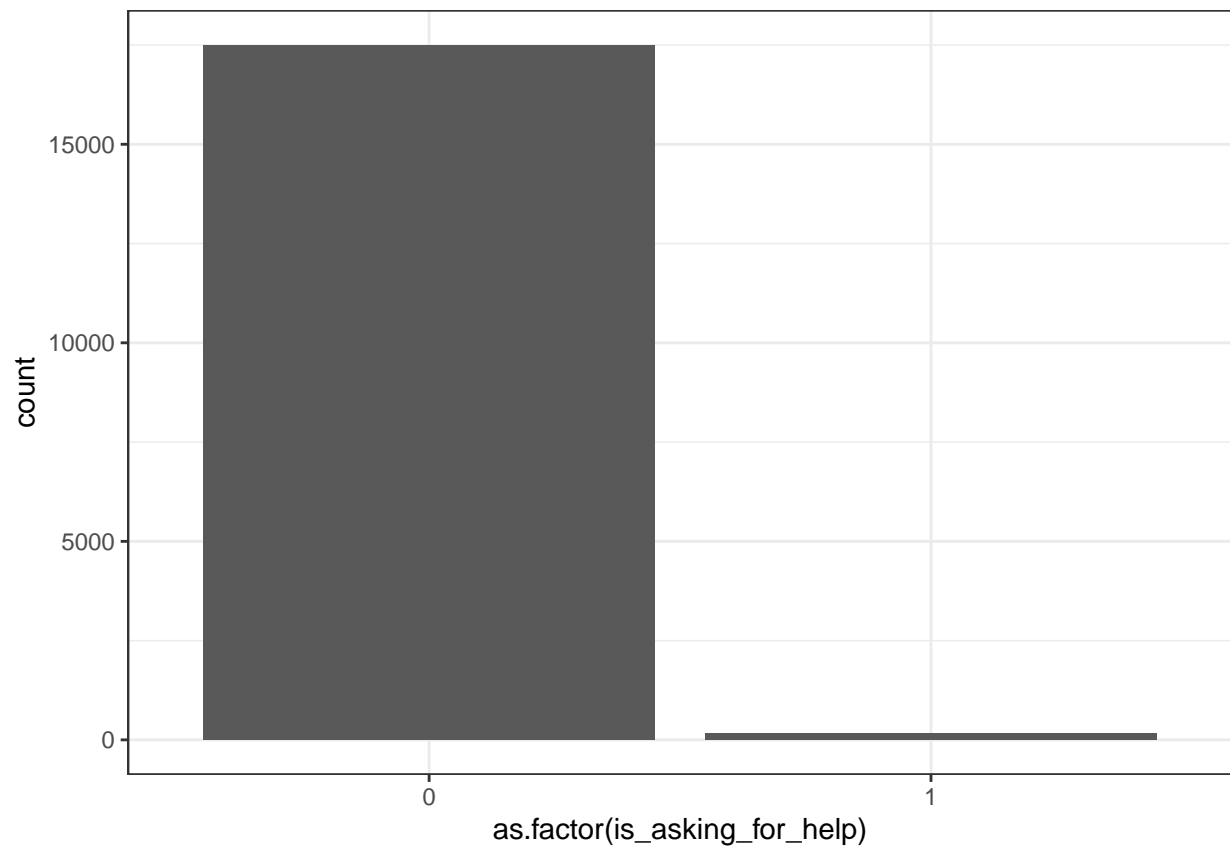
```
ggplot(data=data_clean, aes(x=as.factor(has_one_dollar_reward))) + geom_histogram(stat="count") + theme
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



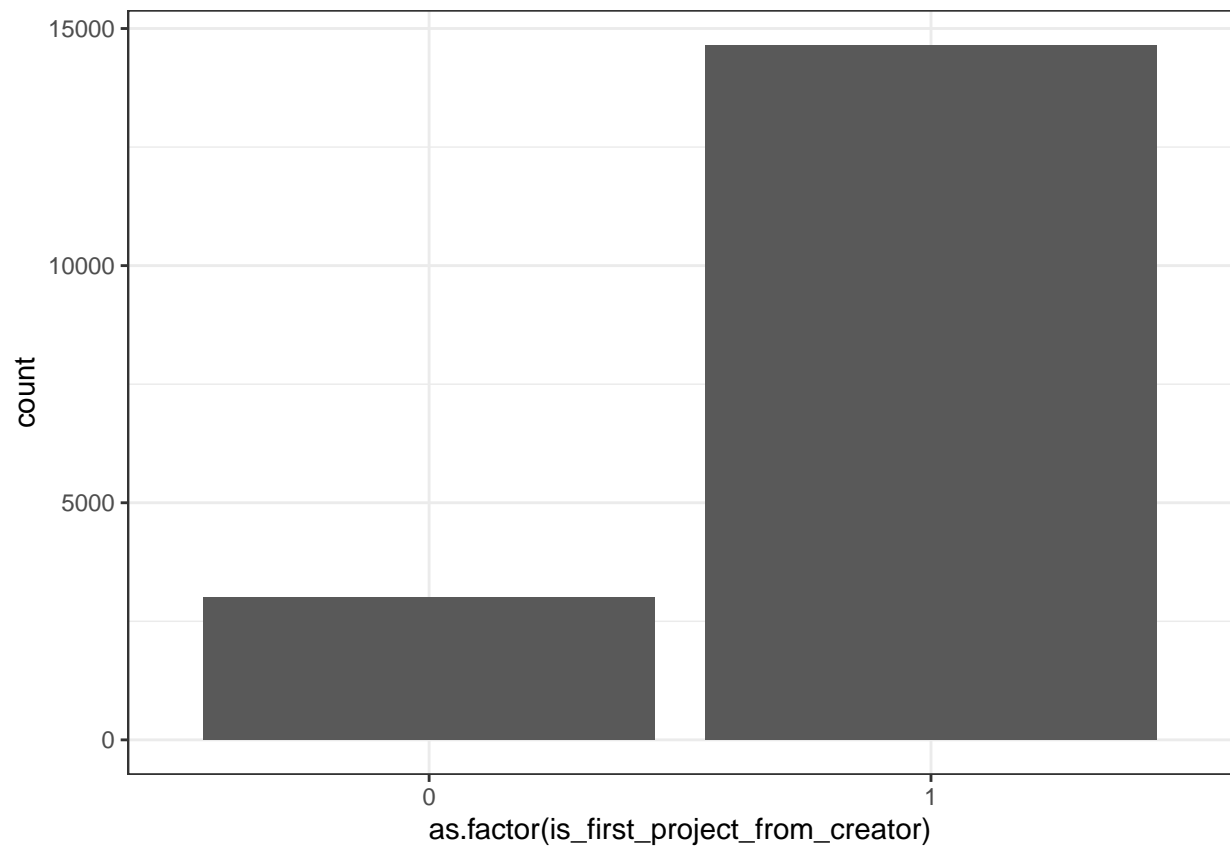
```
ggplot(data=data_clean, aes(x=as.factor(is_asking_for_help))) + geom_histogram(stat="count") + theme_b
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



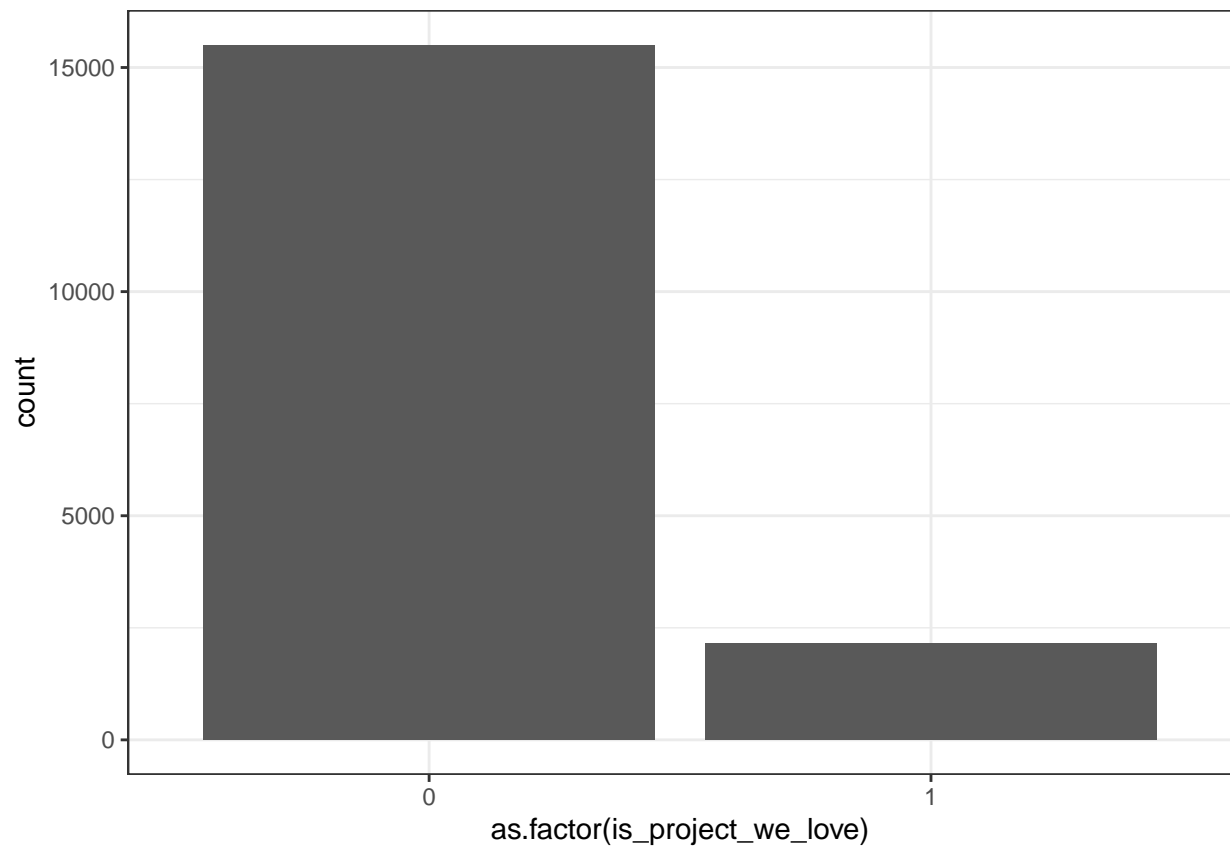
```
ggplot(data=data_clean, aes(x=as.factor(is_first_project_from_creator))) + geom_histogram(stat="count")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

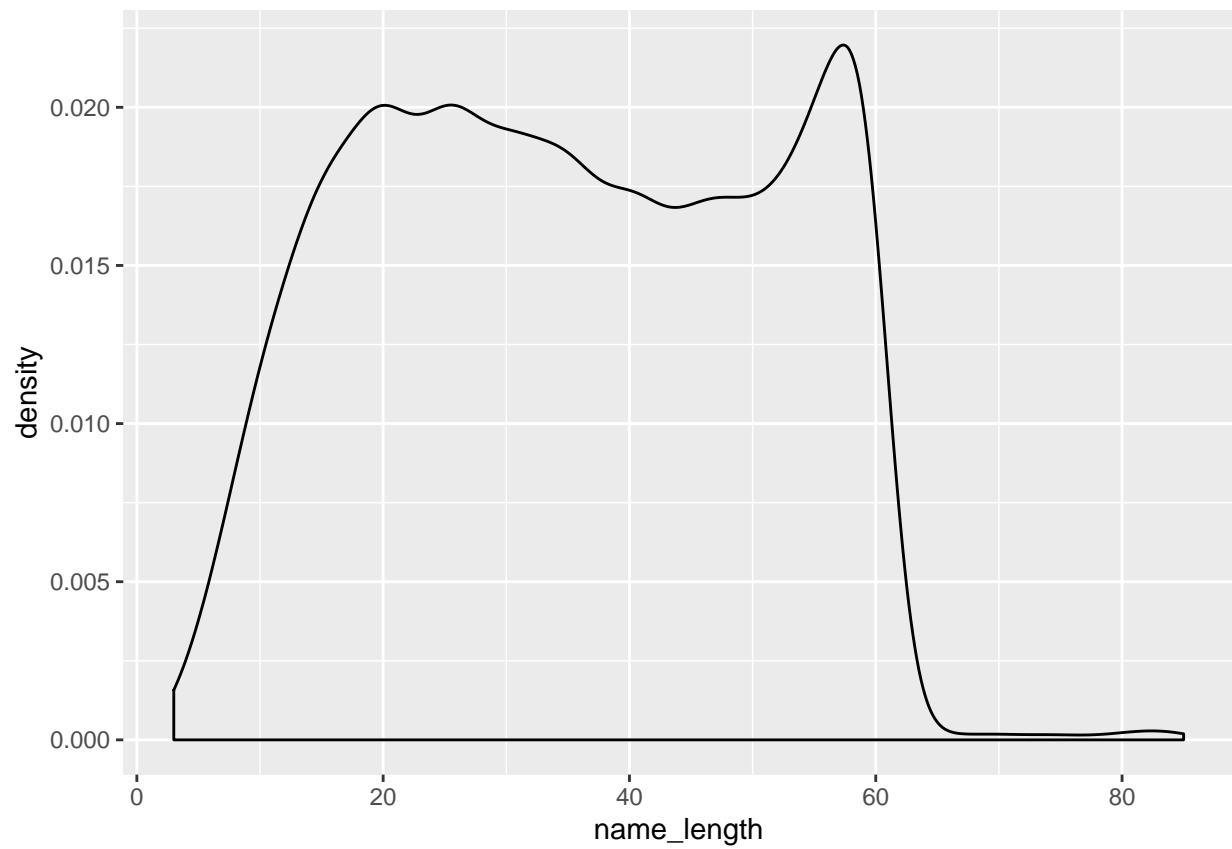


```
ggplot(data=data_clean, aes(x=as.factor(is_project_we_love))) + geom_histogram(stat="count") + theme_b
```

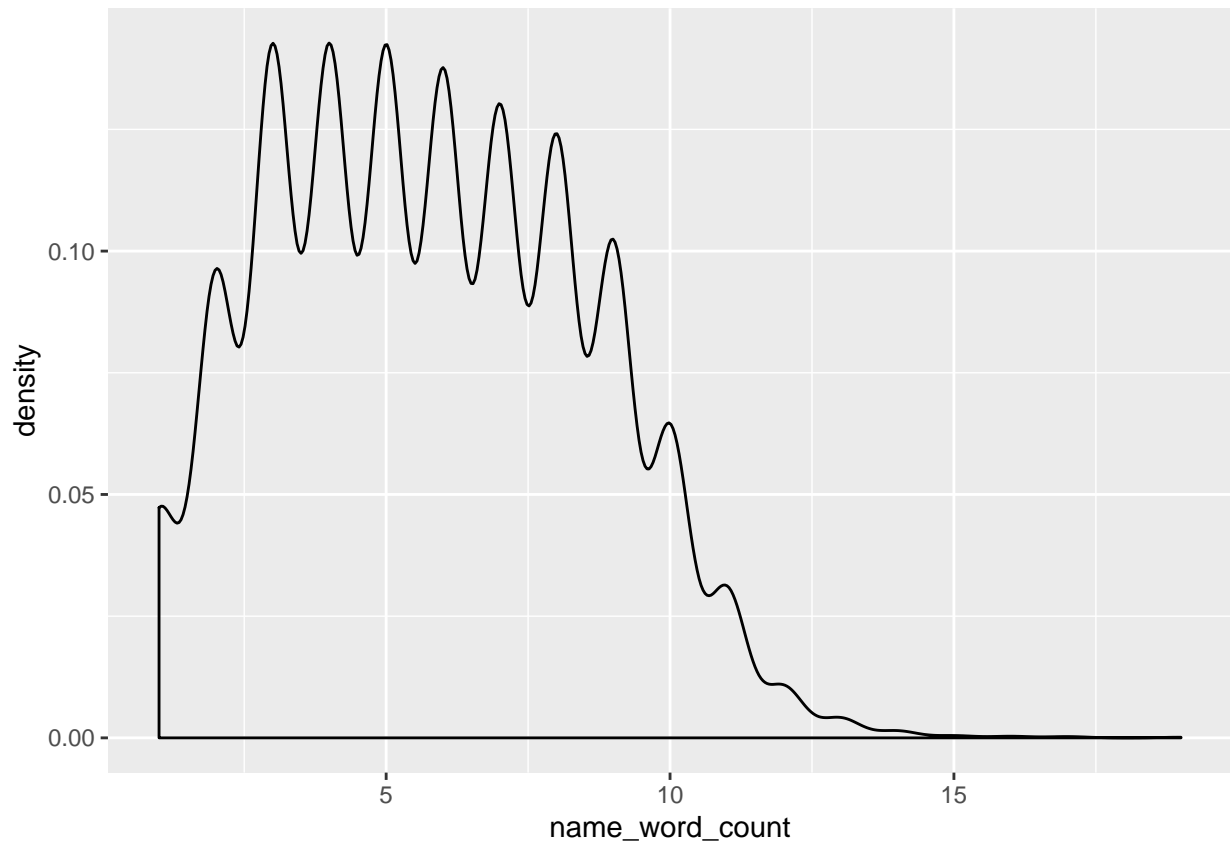
```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



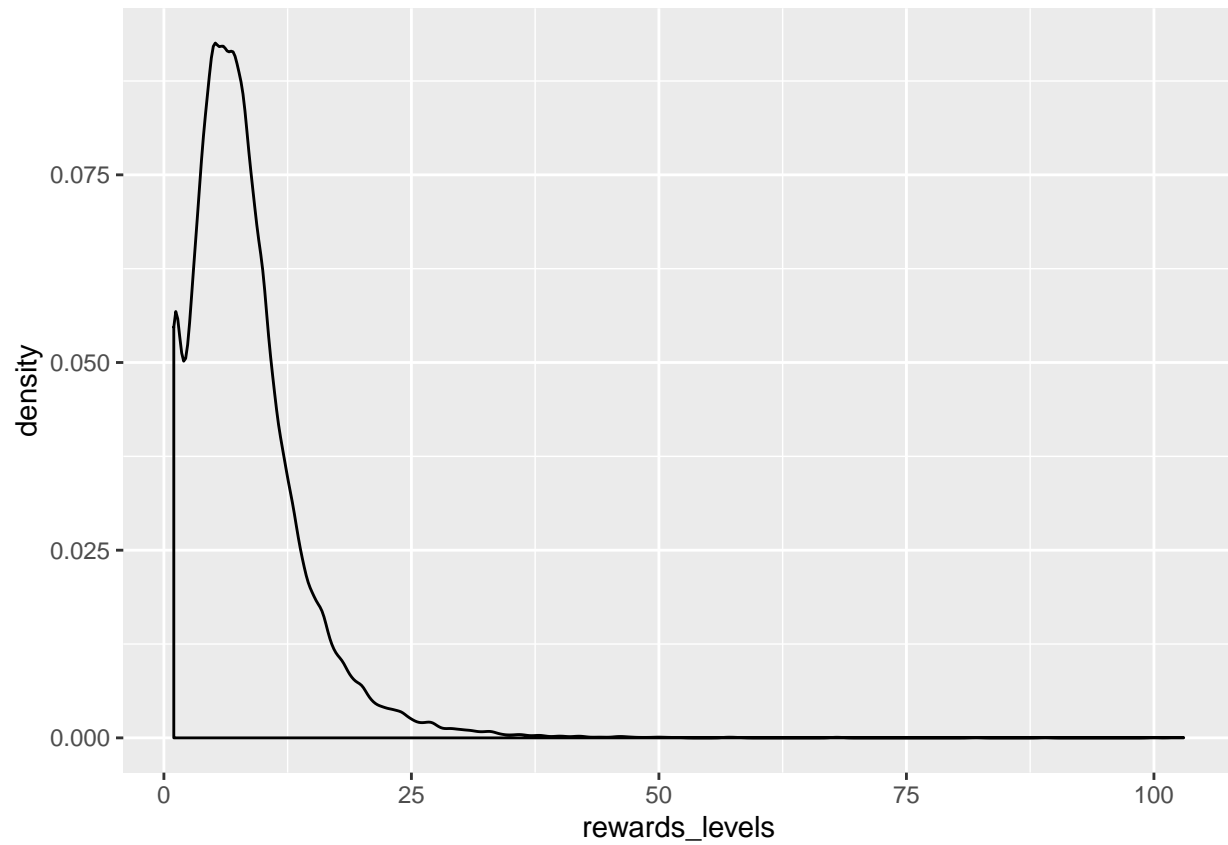
```
ggplot(data=data_clean, aes(x=name_length)) + geom_density(alpha=0.5)
```



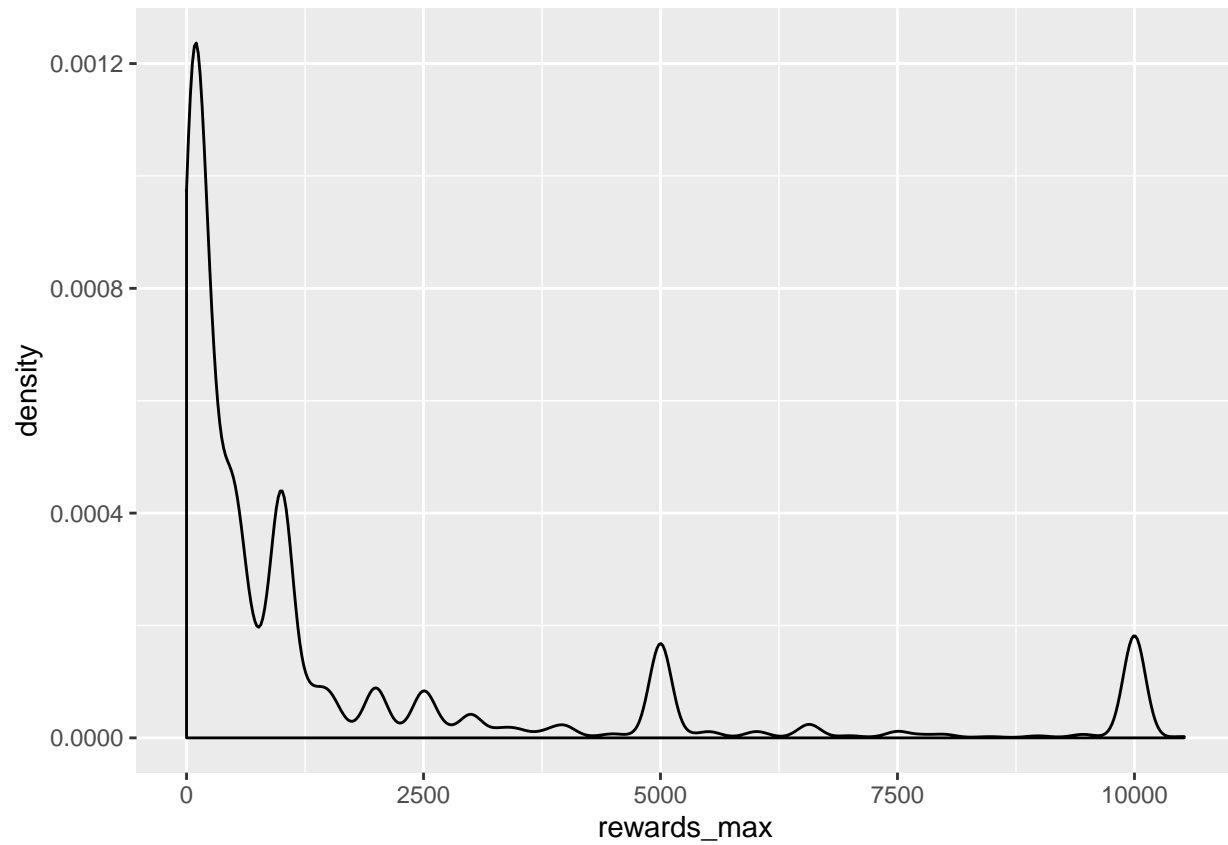
```
ggplot(data=data_clean, aes(x=name_word_count)) + geom_density(alpha=0.5)
```



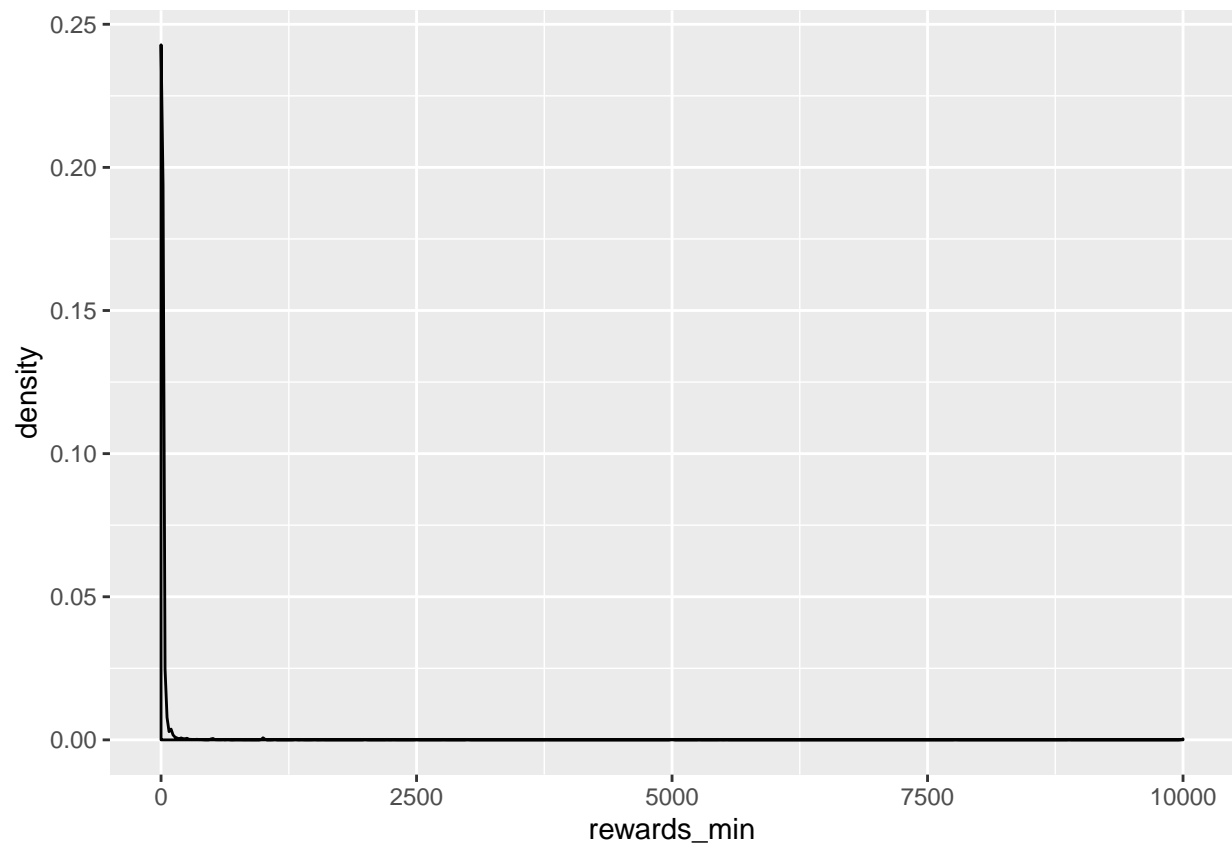
```
ggplot(data=data_clean, aes(x=rewards_levels)) + geom_density(alpha=0.5)
```



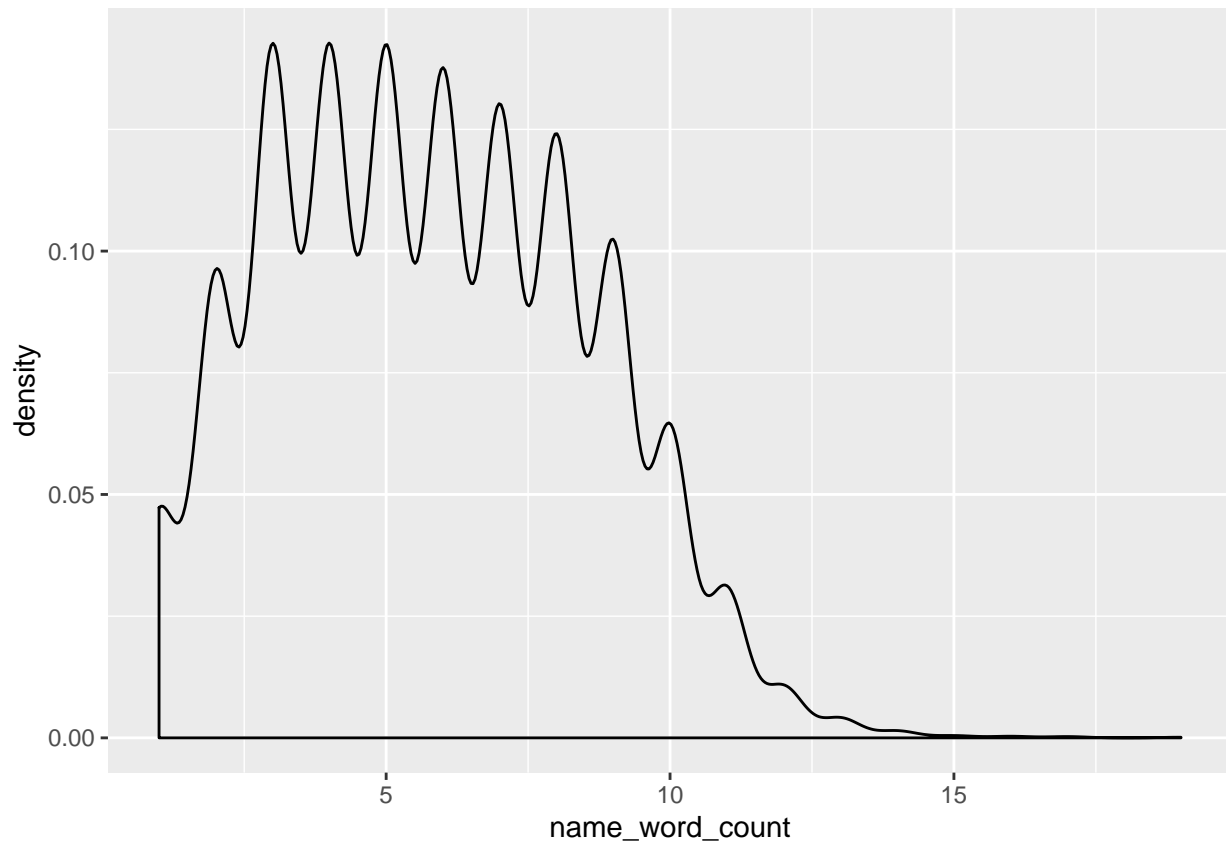
```
ggplot(data=data_clean, aes(x=rewards_max)) + geom_density(alpha=0.5)
```

```
ggplot(data=data_clean, aes(x=rewards_min)) + geom_density(alpha=0.5)
```

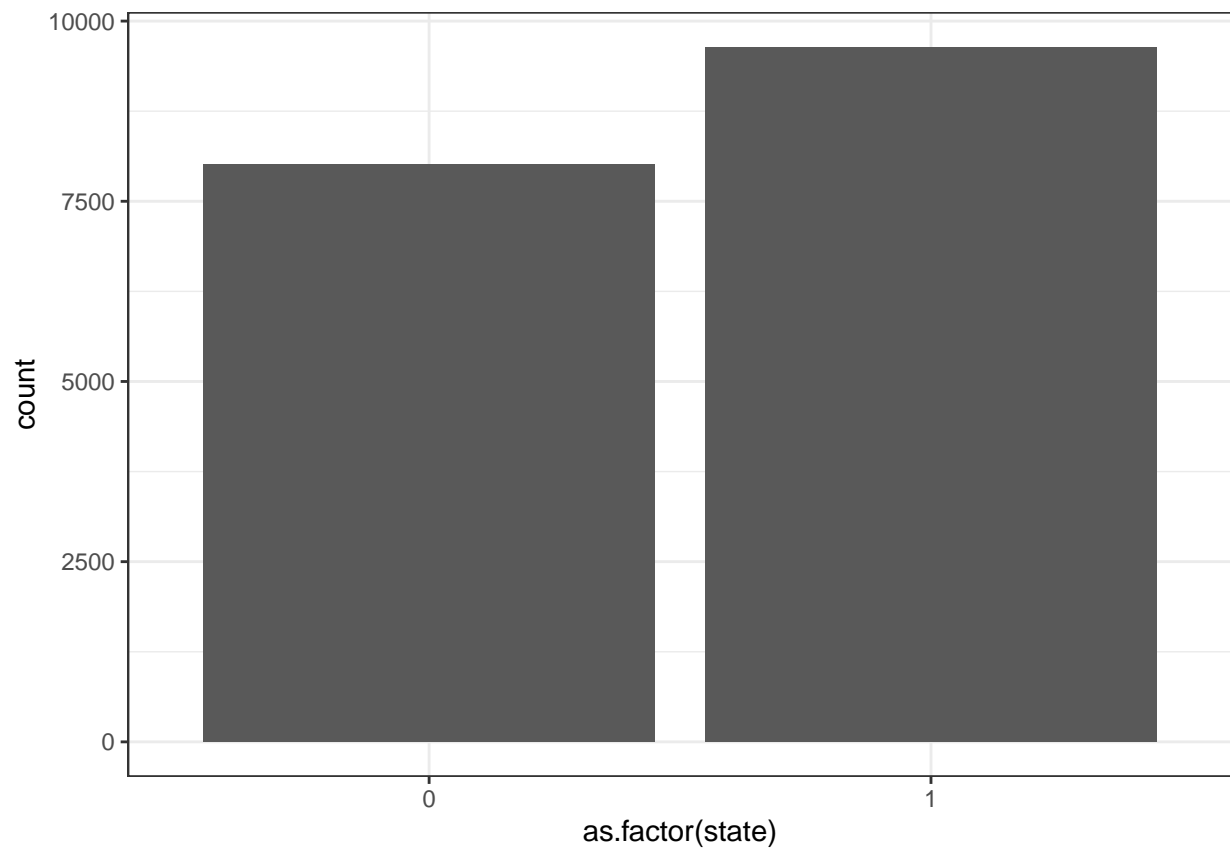


```
ggplot(data=data_clean, aes(x=name_word_count)) + geom_density(alpha=0.5)
```

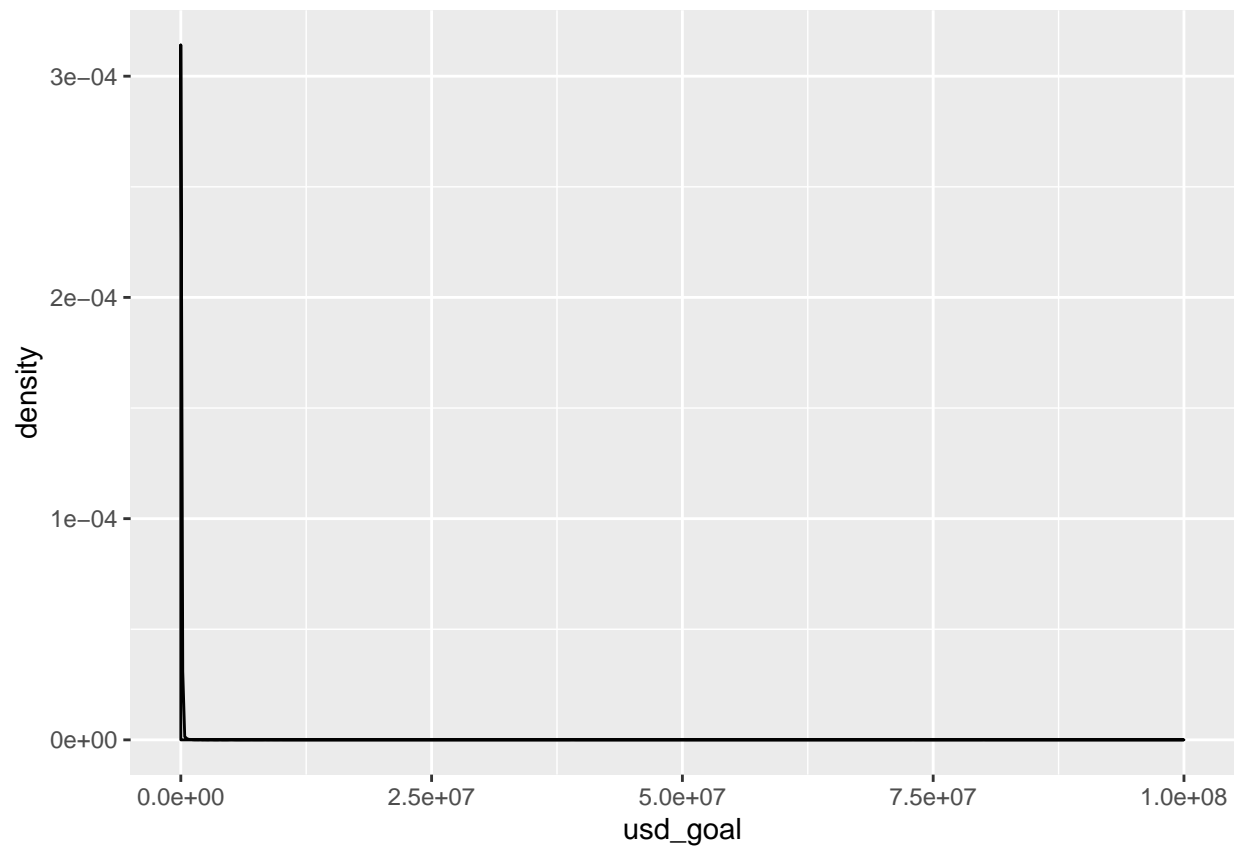


```
ggplot(data=data_clean, aes(x=as.factor(state))) + geom_histogram(stat="count") + theme_bw()
```

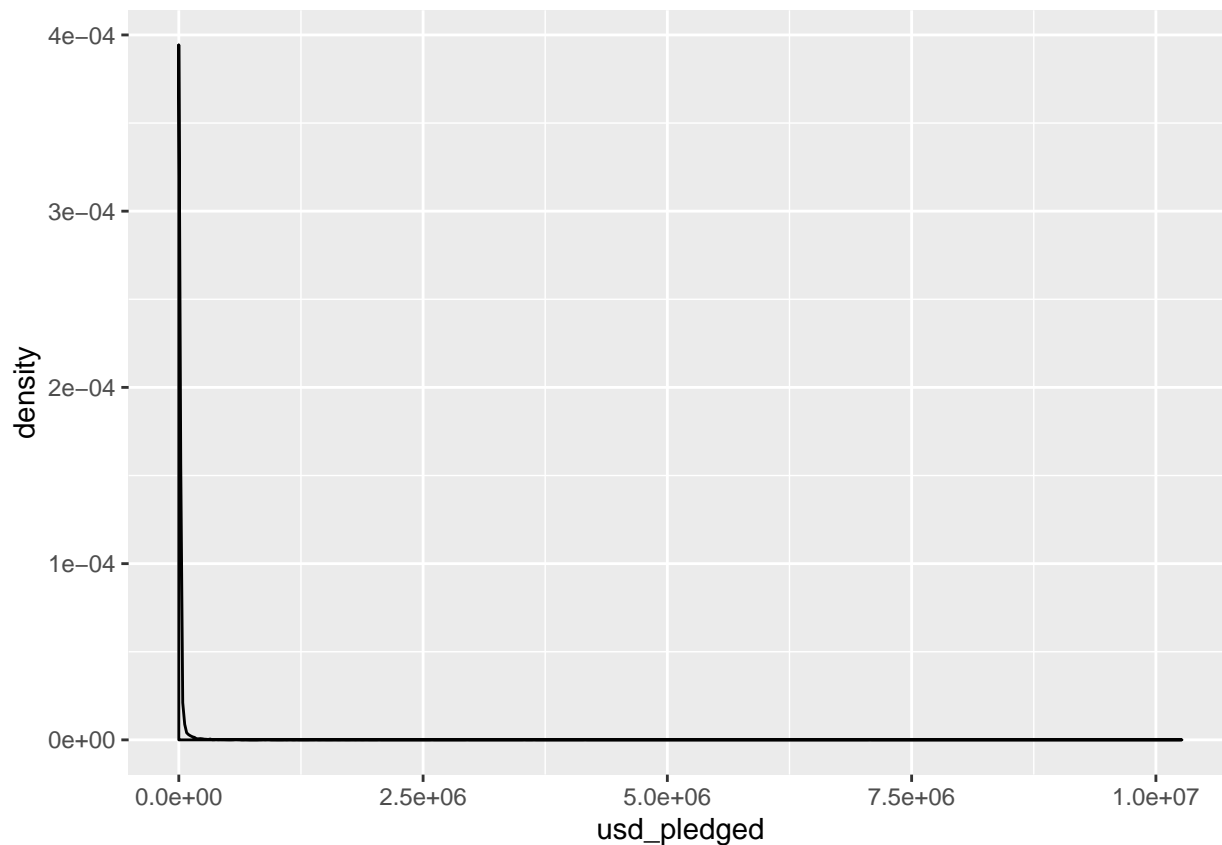
```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



```
ggplot(data=data_clean, aes(x=usd_goal)) + geom_density(alpha=0.5)
```



```
ggplot(data=data_clean, aes(x=usd_pledged)) + geom_density(alpha=0.5)
```



```
colnames(data1)
round(colSums(data1)/nrow(data1),2)

data1 <- data1 %>% mutate(goal_cat_Low_pre=goal_cat_Low+goal_cat_VeryLow) %>%
  mutate(goal_cat_High_pre=goal_cat_Medium+goal_cat_VeryHigh) %>%

goal_cats <- c("goal_cat_Low","goal_cat_Medium","goal_cat_VeryHigh","goal_cat_VeryLow")

data2 <- read.csv("../data/Kickstarter_2019-11-14T03_20_27_004Z/Kickstarter054.csv", sep=",")

data3 <- read.csv("../data/dspro.csv", sep=",")

data3 <- data3 %>% mutate(duration=as.Date(deadline)-as.Date(launched)) %>%
  filter(duration<1000) %>%
  filter(usd_goal_real<250000)

ggplot(data=data3, aes(x=as.numeric(duration))) + geom_density(alpha=0.5)
ggplot(data=data3, aes(x=usd_goal_real)) + geom_density(alpha=0.5)
```