Stones

Nick Lauerman

12/3/2019

Contents

Libraries

```
library(MASS)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
## <U+2713> ggplot2 3.2.1     <U+2713> purrr   0.3.3
## <U+2713> tibble  2.1.3     <U+2713> dplyr   0.8.3
## <U+2713> tidyr   1.0.0     <U+2713> stringr 1.4.0
## <U+2713> readr   1.3.1     <U+2713> forcats 0.4.0
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x dplyr::select() masks MASS::select()
```

read data

```
stones <- read_csv(file = "./data/stones.csv")
```

```
## Parsed with column specification:
## cols(
##   mass = col_double(),
##   stone = col_character(),
##   `price per gram` = col_double()
## )
```

```
stones$ct <- stones$mass * 5
stones$cost_raw <- stones$mass * stones$`price per gram`
stones$stone <- factor(stones$stone)
```

summary stats

```
summary(stones$ct)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.70    9.10   12.95   21.22   19.65  215.00
```

```
sd(stones$ct)
```

```
## [1] 28.14101
```

```
summary(stones$cost_raw)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00826 0.05180 0.07308 0.22551 0.11086 4.73000
```

```
sd(stones$cost_raw)
```

```
## [1] 0.5762471
```

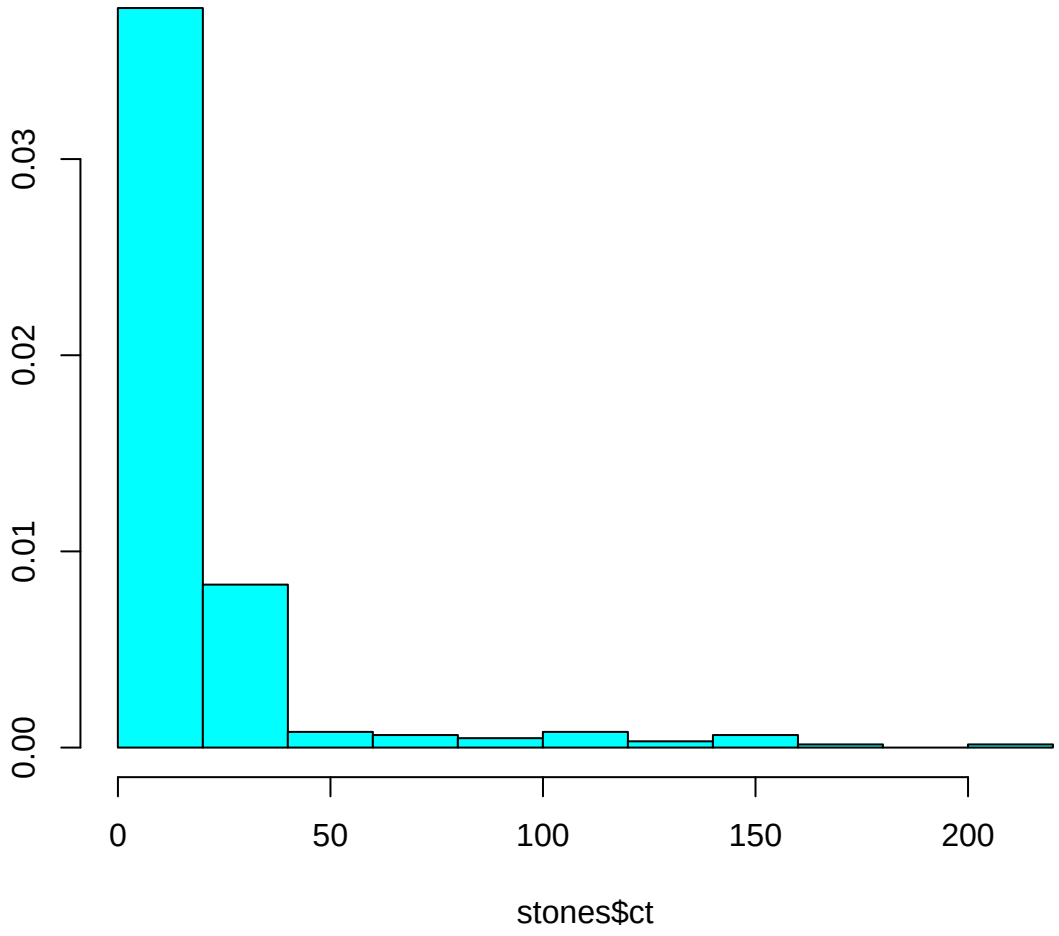
```
sumStats <- stones %>%
  group_by(stone) %>%
  summarise(count = n(),
            avgSize = mean(ct),
            sdSize = sd(ct),
            medSize = median(ct),
            largest = max(ct),
            smallest = min(ct),
            cost = mean(cost_raw))

sumStats
```

```
## # A tibble: 4 x 8
##   stone    count avgSize sdSize medSize largest smallest   cost
##   <fct>    <int>   <dbl>  <dbl>   <dbl>   <dbl>    <dbl>  <dbl>
## 1 amethyst   162    11.1   4.27    10.1    25       3.85 0.0620
## 2 Emerald     46    29.0  38.7     14.8   178.      0.7  0.342
## 3 quartz      91    20.6   9.37    19.2    69.4     3.55 0.0945
## 4 Topaz       14   118.   38.9    113.    215      63.8  2.59
```

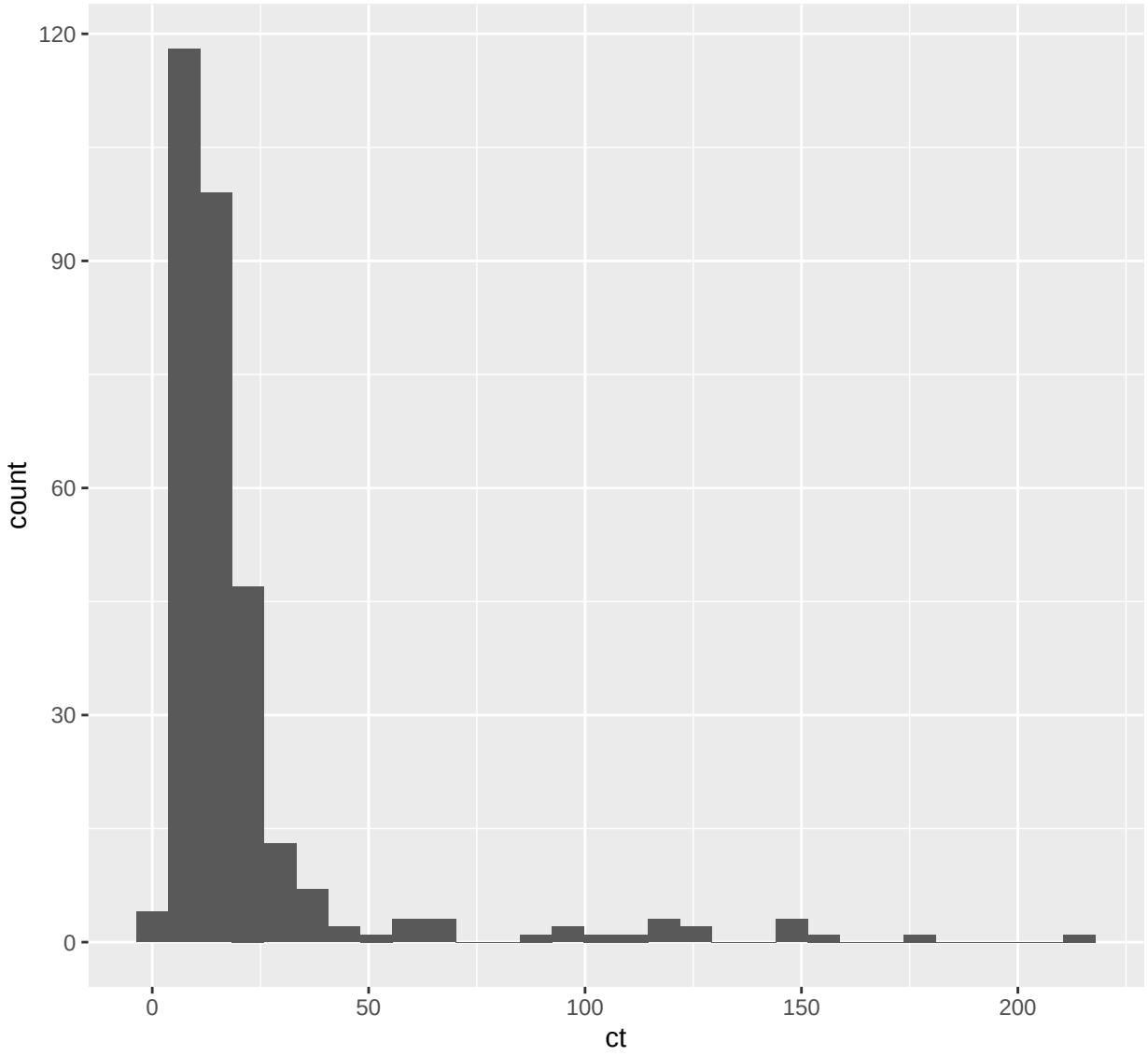
plots

```
truehist(stones$ct)
```

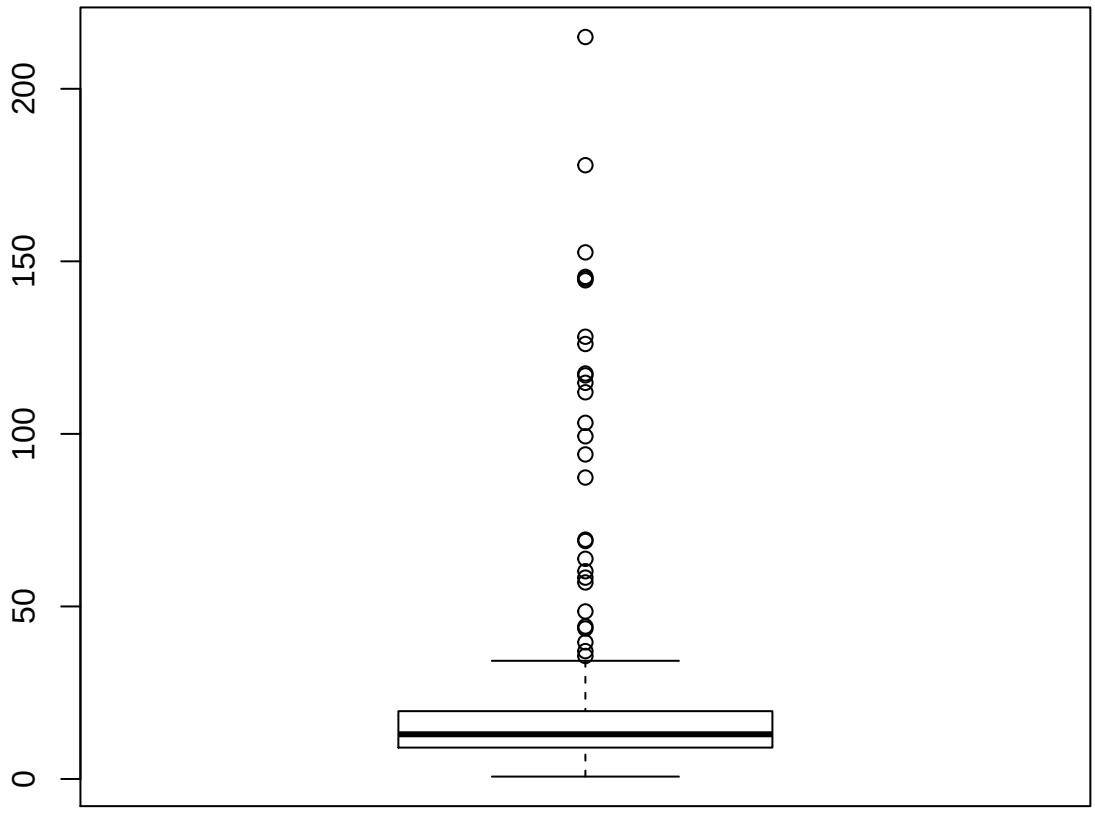



```
ggplot(data = stones,
       aes(ct)) +
     geom_histogram()
```

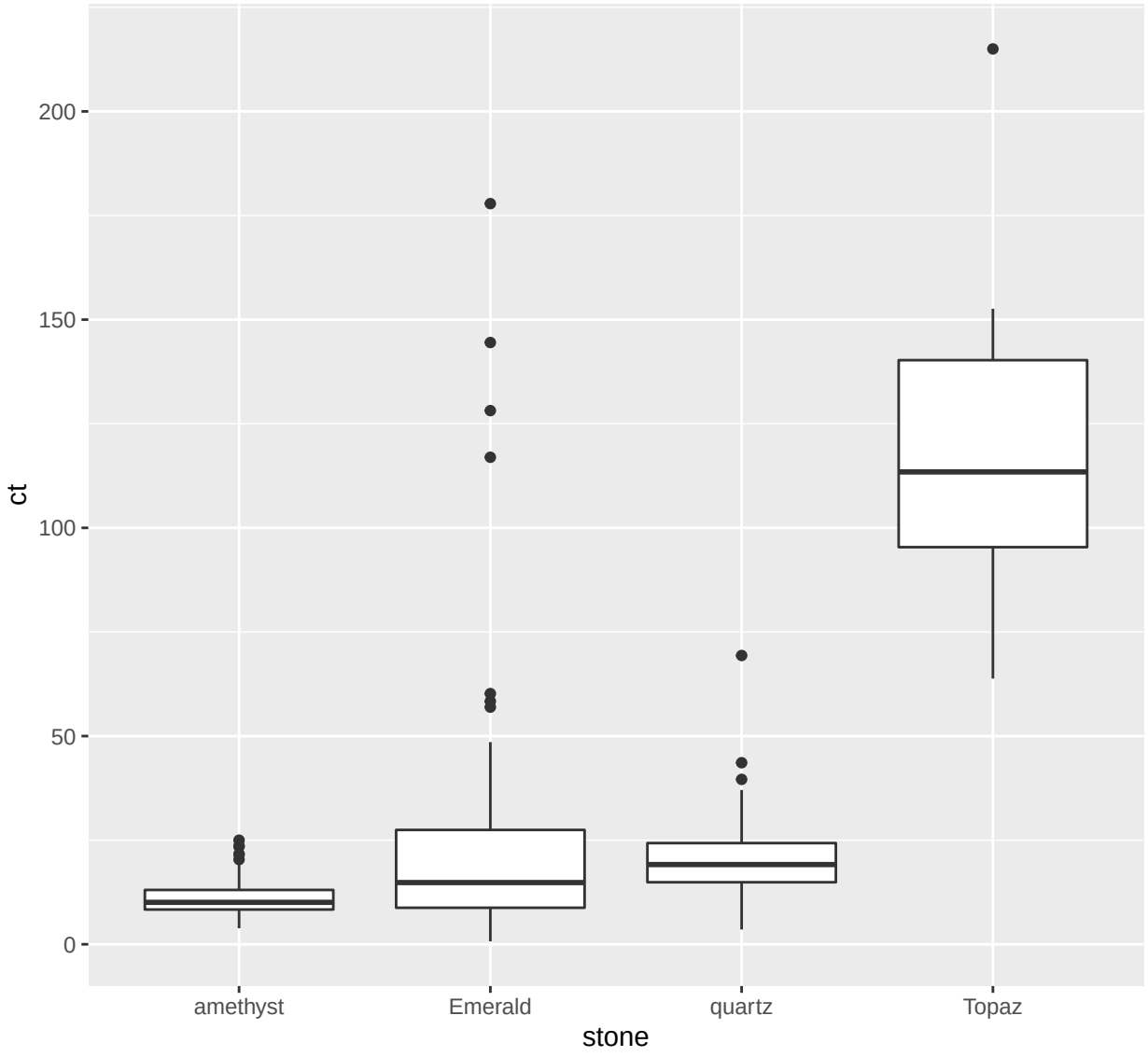
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



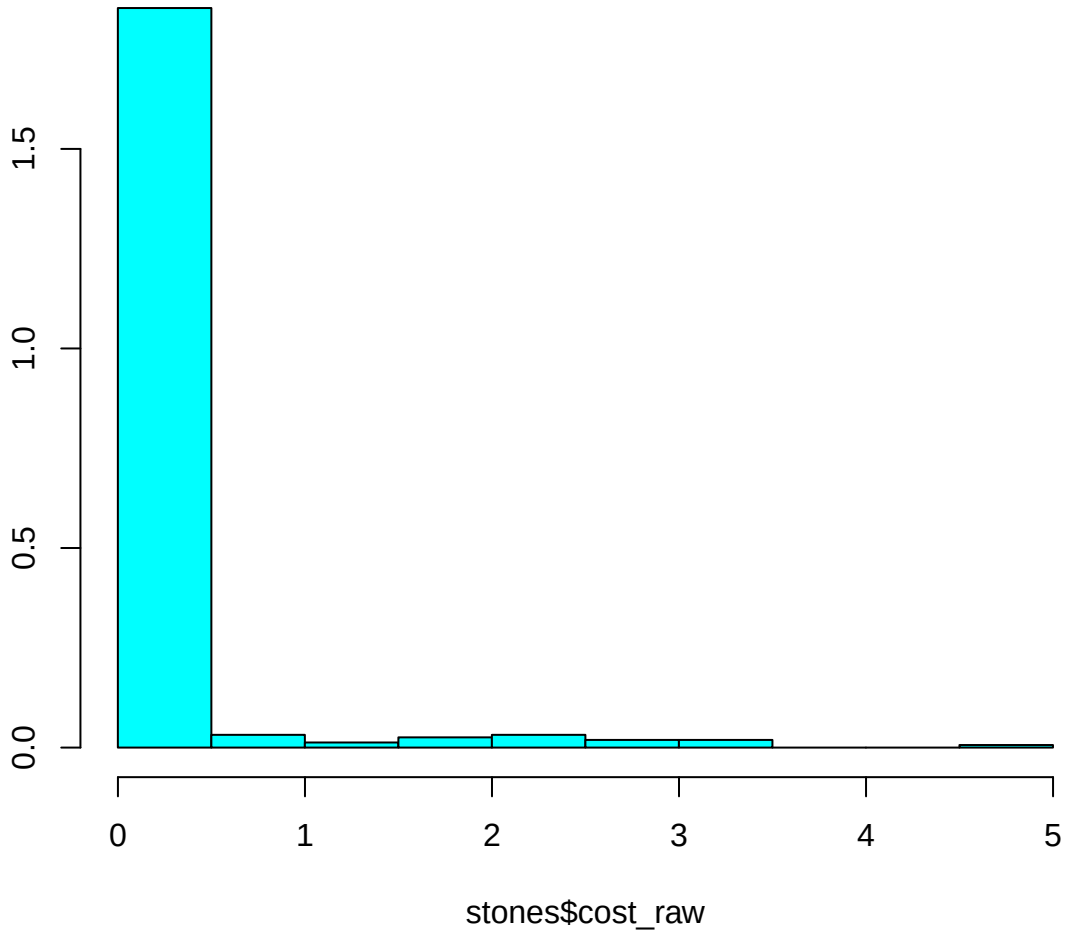

```
boxplot(stones$ct)
```




```
ggplot(data = stones,
       aes(y = ct,
           x = stone)) +
     geom_boxplot()
```

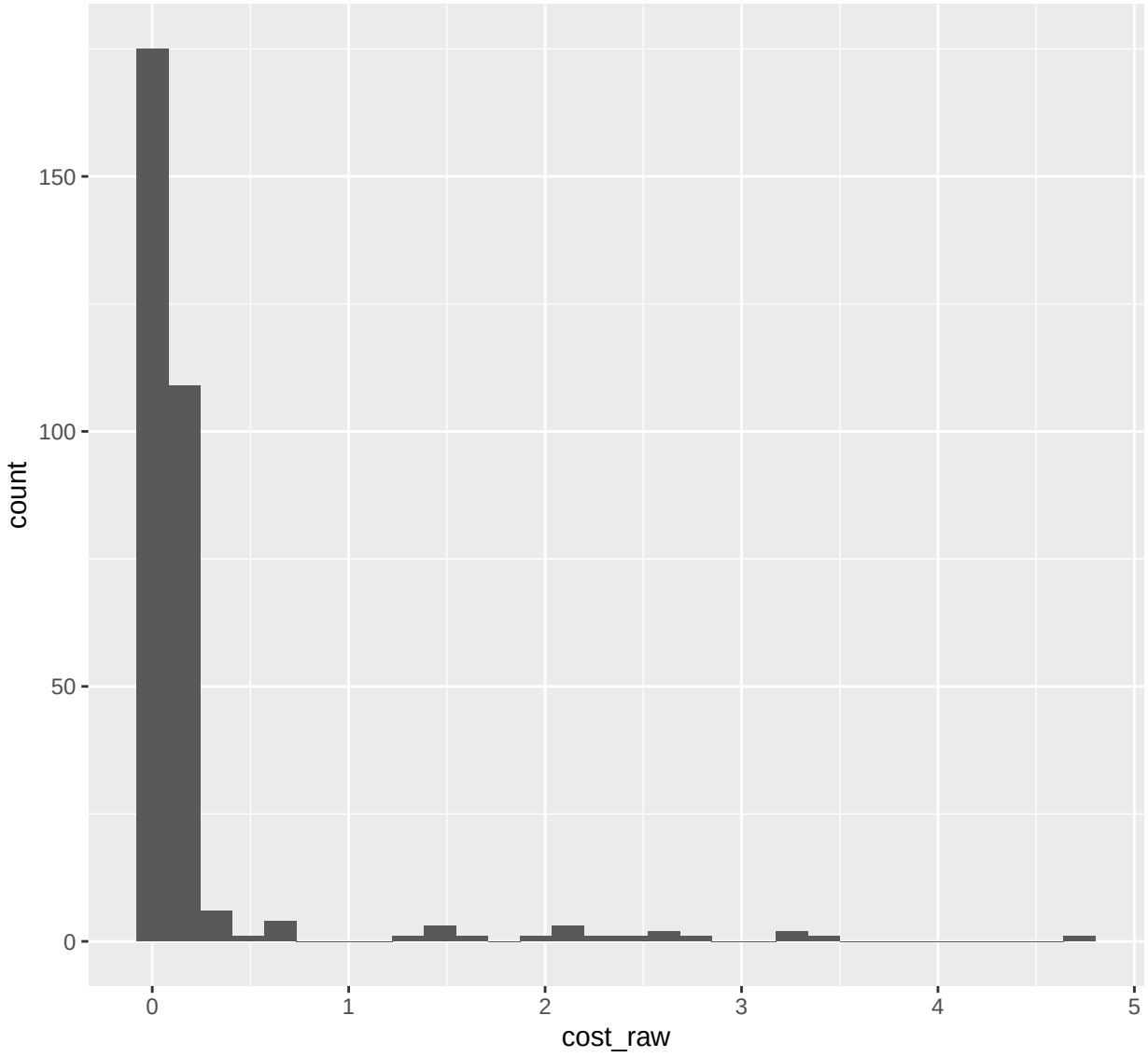



```
truehist(stones$cost_raw)
```



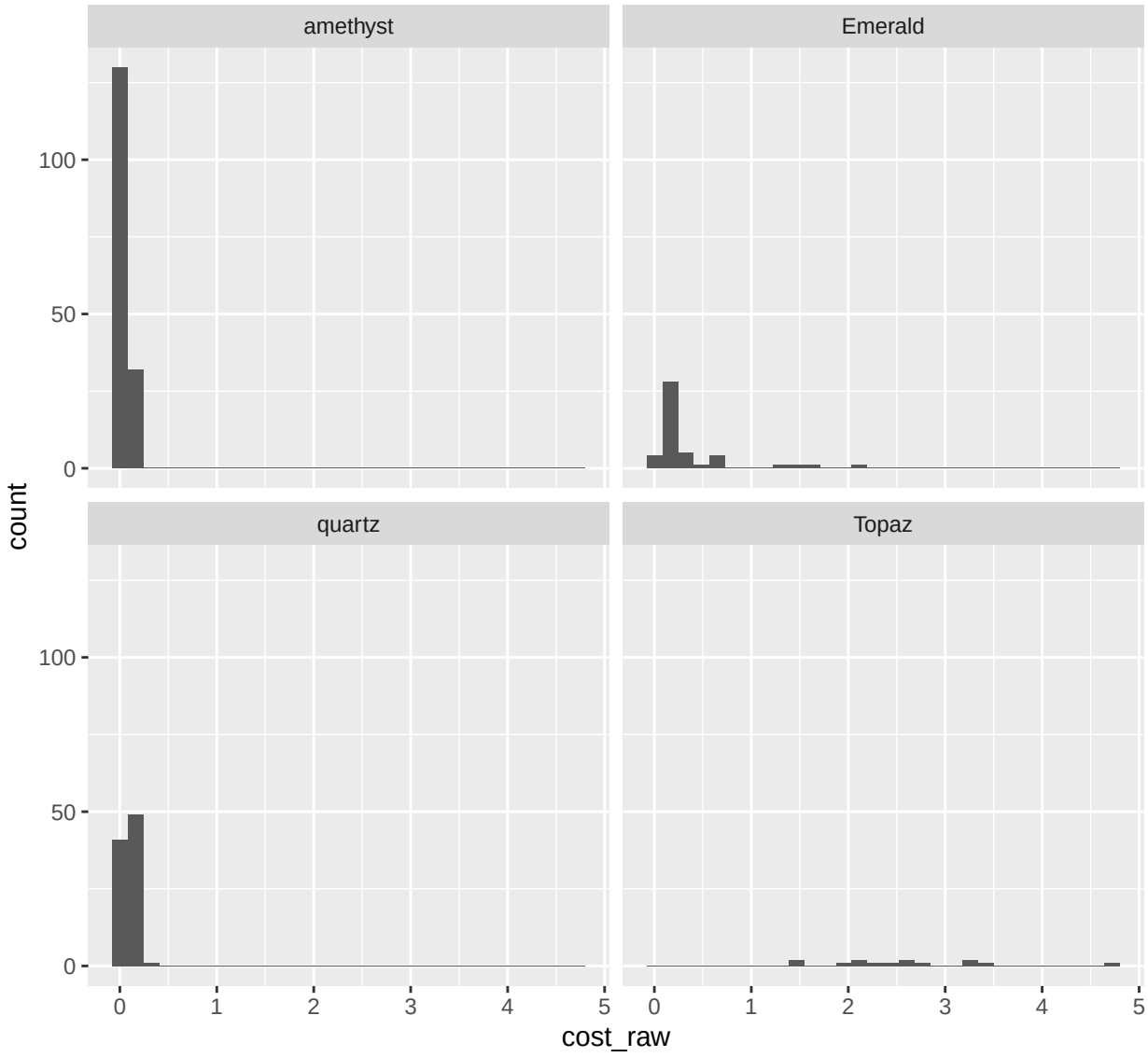

```
ggplot(data = stones,
       aes(cost_raw)) +
     geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

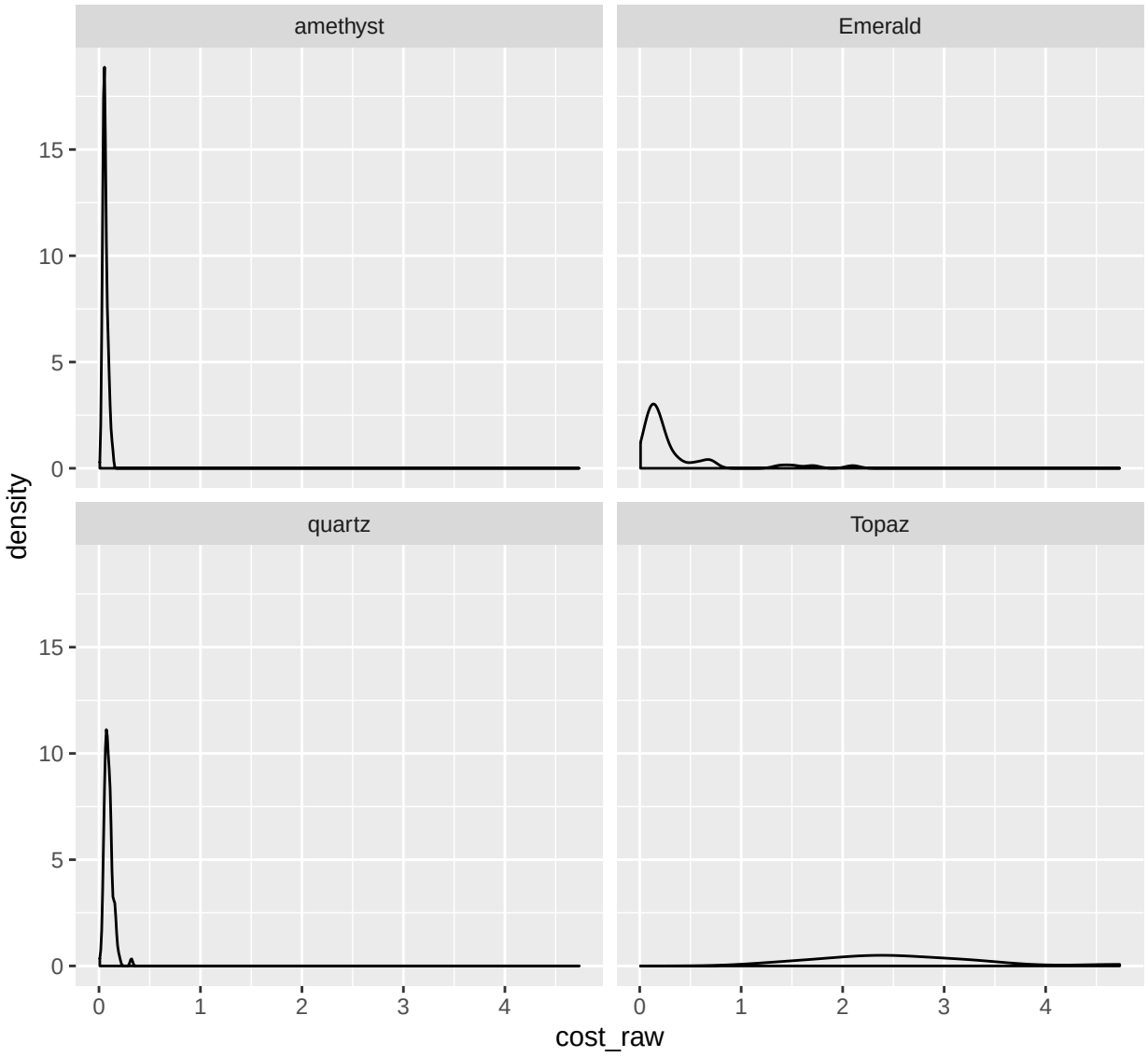



```
ggplot(data = stones,
       aes(cost_raw)) +
     geom_histogram() +
  facet_wrap( ~ stone)
```

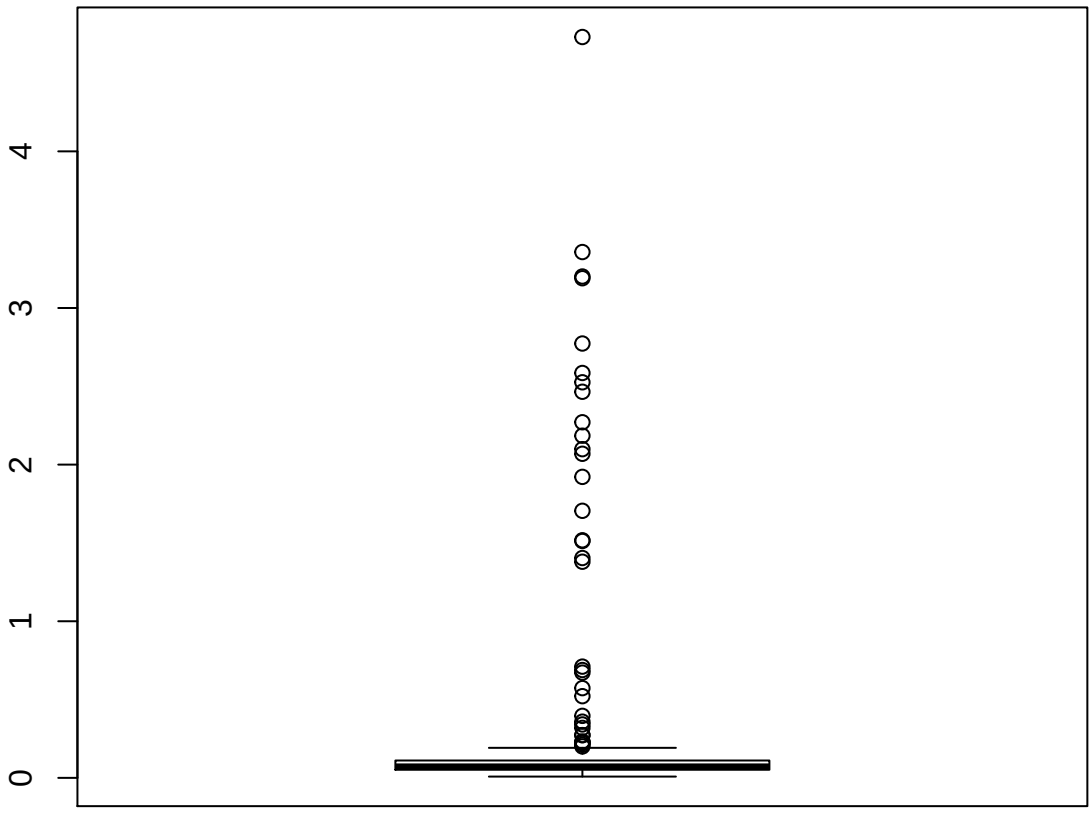
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



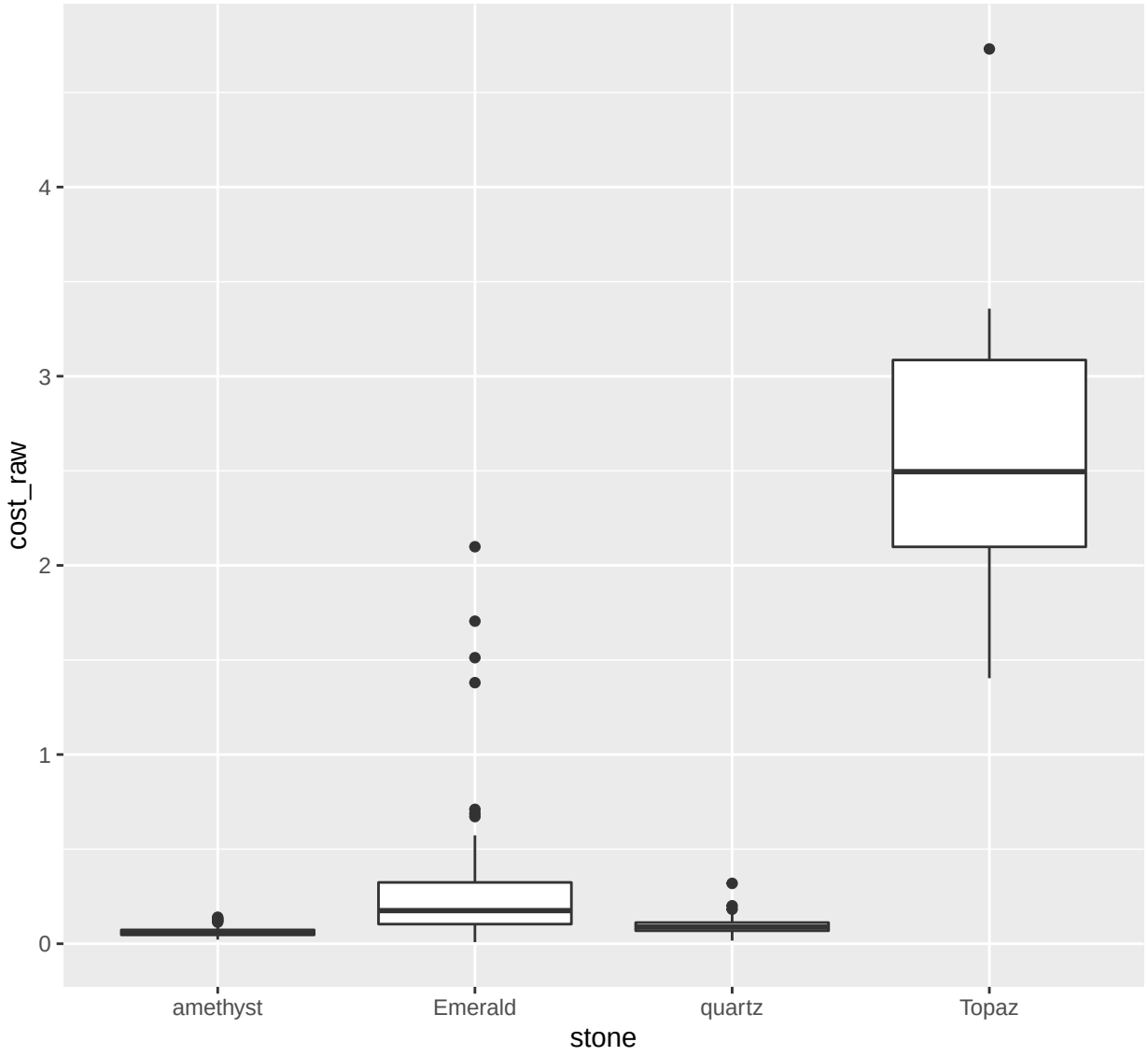

```
ggplot(data = stones,
       aes(cost_raw)) +
     geom_density() +
  facet_wrap( ~ stone)
```




```
boxplot(stones$cost_raw)
```




```
ggplot(data = stones,
       aes(y = cost_raw,
           x = stone)) +
     geom_boxplot()
```


cost_raw
4
3
2
1
0
amethyst
Emerald
quartz
Topaz
stone