Intro to Grammar of Graphics (ggplot2)

Jason Bryer

September 9, 2014

ggplot2

`ggplot2` (Wickham, 2009) is an R package that implements the *Graammar of Graphics* introduced by Wilkinson (2005).

(Journal of Statistical Software paper)[`http://www.jstatsoft.org/v17/b03/paper`]

The Lego Package

```
# https://github.com/seankross/lego
devtools::install_github("seankross/lego")
```

```
library(lego)
library(psych)
```

The Lego Package

```
data(legosets)
head(legosets, n=3)
```

```
##   Item_Number                      Name Year           Theme
## 1       10241 Maersk Line Triple-E 2014 Advanced Models
## 2       10242   Mini Cooper MK VII 2014 Advanced Models
## 3       10243  Parisian Restaurant 2014 Advanced Models
##            Subtheme Pieces Minifigures
## 1            Maersk   1518          NA
## 2          Vehicles   1077          NA
## 3 Modular Buildings   2469           5
##                                                Image_URL
## 1 http://www.1000steine.com/brickset/images/10241-1.jpg
## 2 http://www.1000steine.com/brickset/images/10242-1.jpg
## 3 http://www.1000steine.com/brickset/images/10243-1.jpg
##   USD_MSRP CAD_MSRP EUR_MSRP Packaging   Availability
## 1   149.99      180   129.99       Box LEGO exclusive
## 2    99.99      120    89.99       Box LEGO exclusive
```

Parts of a ggplot2 statement

- Data
  ggplot(myDataFrame, aes(x=x, y=y)

Parts of a ggplot2 statement

- Data
  ggplot(myDataFrame, aes(x=x, y=y)
- Layers
  geom_point(), geom_histogram()

Parts of a ggplot2 statement

- Data
  ggplot(myDataFrame, aes(x=x, y=y)
- Layers
  geom_point(), geom_histogram()
- Facets
  facet_wrap(~ cut), facet_grid(~ cut)

Parts of a ggplot2 statement

- Data
  ggplot(myDataFrame, aes(x=x, y=y)
- Layers
  geom_point(), geom_histogram()
- Facets
  facet_wrap(~ cut), facet_grid(~ cut)
- Scales
  scale_y_log10()

Parts of a ggplot2 statement

- Data
  ggplot(myDataFrame, aes(x=x, y=y)
- Layers
  geom_point(), geom_histogram()
- Facets
  facet_wrap(~ cut), facet_grid(~ cut)
- Scales
  scale_y_log10()
- Other options
  ggtitle('my title'), ylim(c(0, 10000)), xlab('x-axis label')

There are Lots of Geoms

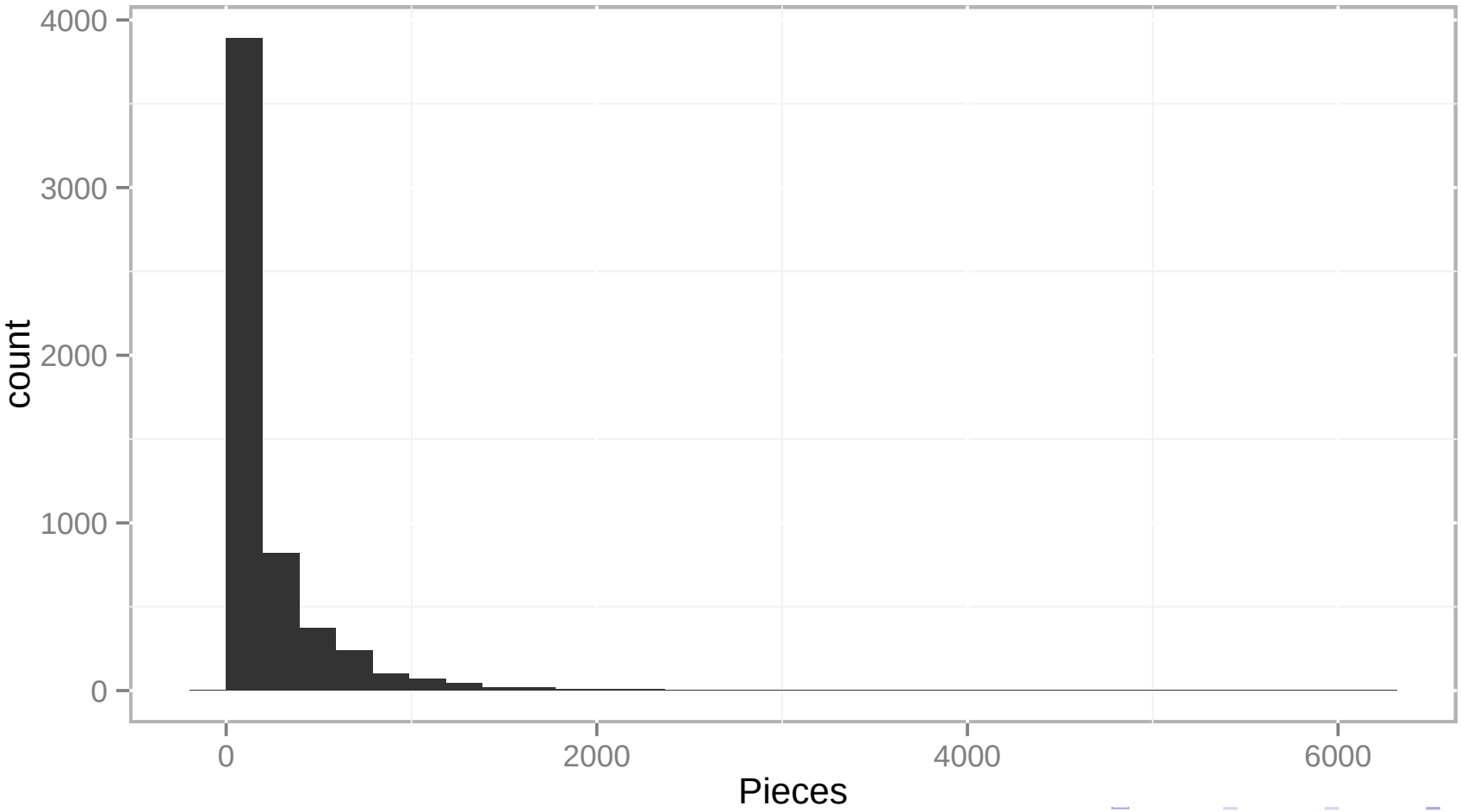
```
ls("package:ggplot2")[substr(ls("package:ggplot2"),
                             1, 5) == 'geom_']
```

```
##  [1] "geom_abline"     "geom_area"       "geom_bar"
##  [4] "geom_bin2d"      "geom_blank"      "geom_boxplot"
##  [7] "geom_contour"    "geom_crossbar"   "geom_density"
## [10] "geom_density2d"  "geom_dotplot"    "geom_errorbar"
## [13] "geom_errorbarh"  "geom_freqpoly"   "geom_hex"
## [16] "geom_histogram"  "geom_hline"      "geom_jitter"
## [19] "geom_line"       "geom_linerange"  "geom_map"
## [22] "geom_path"       "geom_point"      "geom_pointrang
## [25] "geom_polygon"    "geom_quantile"   "geom_raster"
## [28] "geom_rect"       "geom_ribbon"     "geom_rug"
## [31] "geom_segment"    "geom_smooth"     "geom_step"
## [34] "geom_text"       "geom_tile"       "geom_violin"
## [37] "geom_vline"
```

Histograms

```
ggplot(legosets, aes(x=Pieces)) +
    geom_histogram()
```

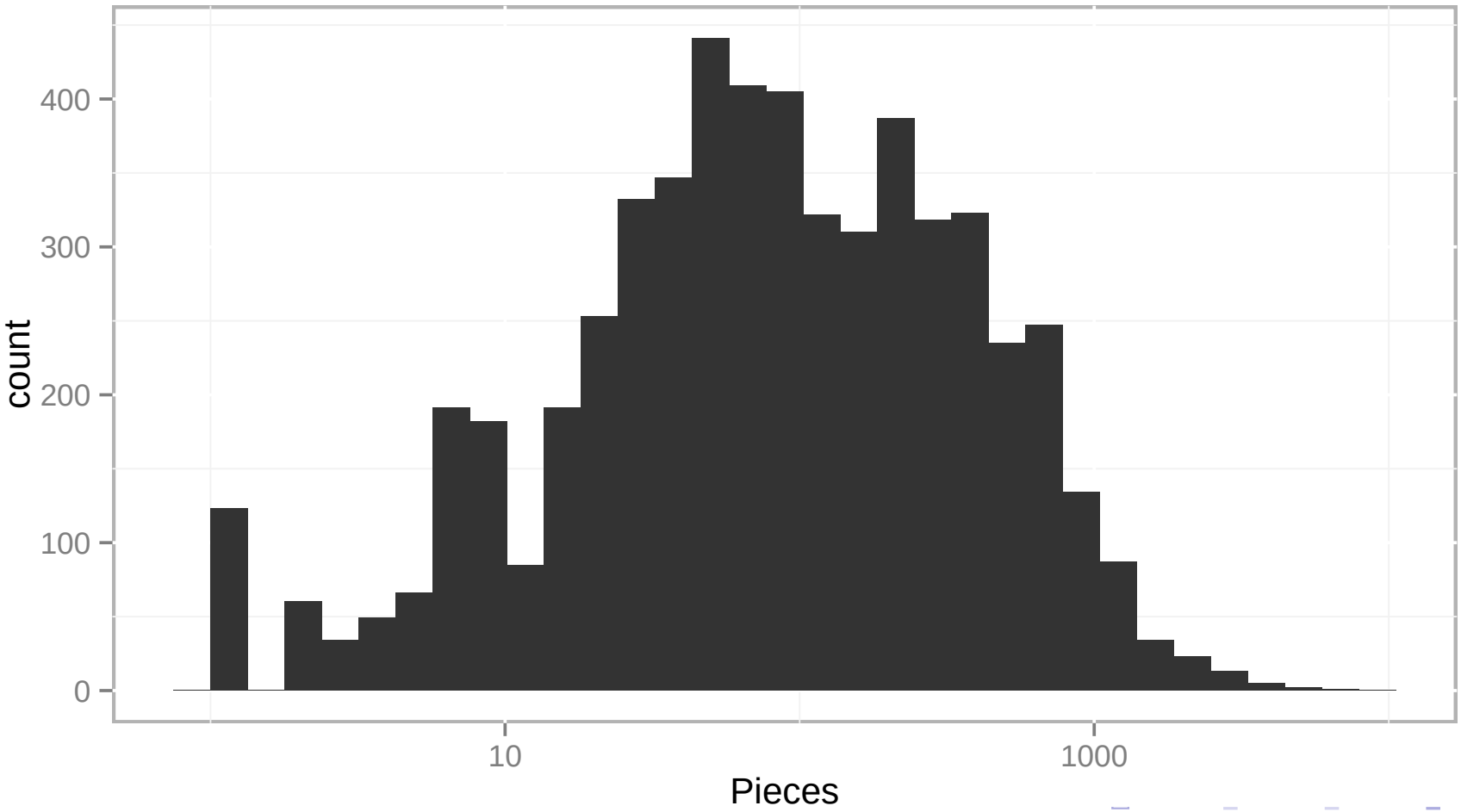
```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth
```



Log Transformations

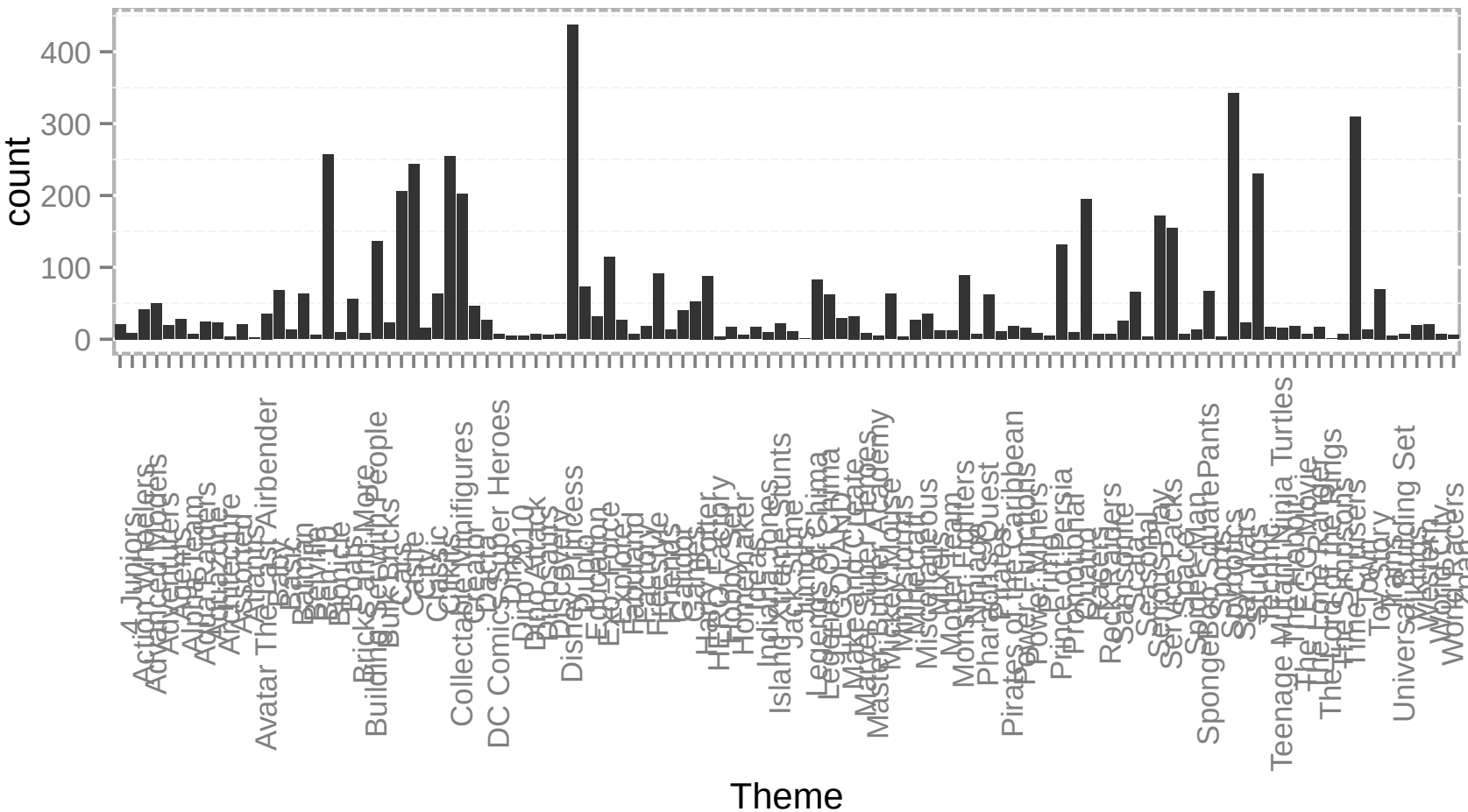
```
ggplot(legosets, aes(x=Pieces)) +
    geom_histogram() + scale_x_log10()
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth
```



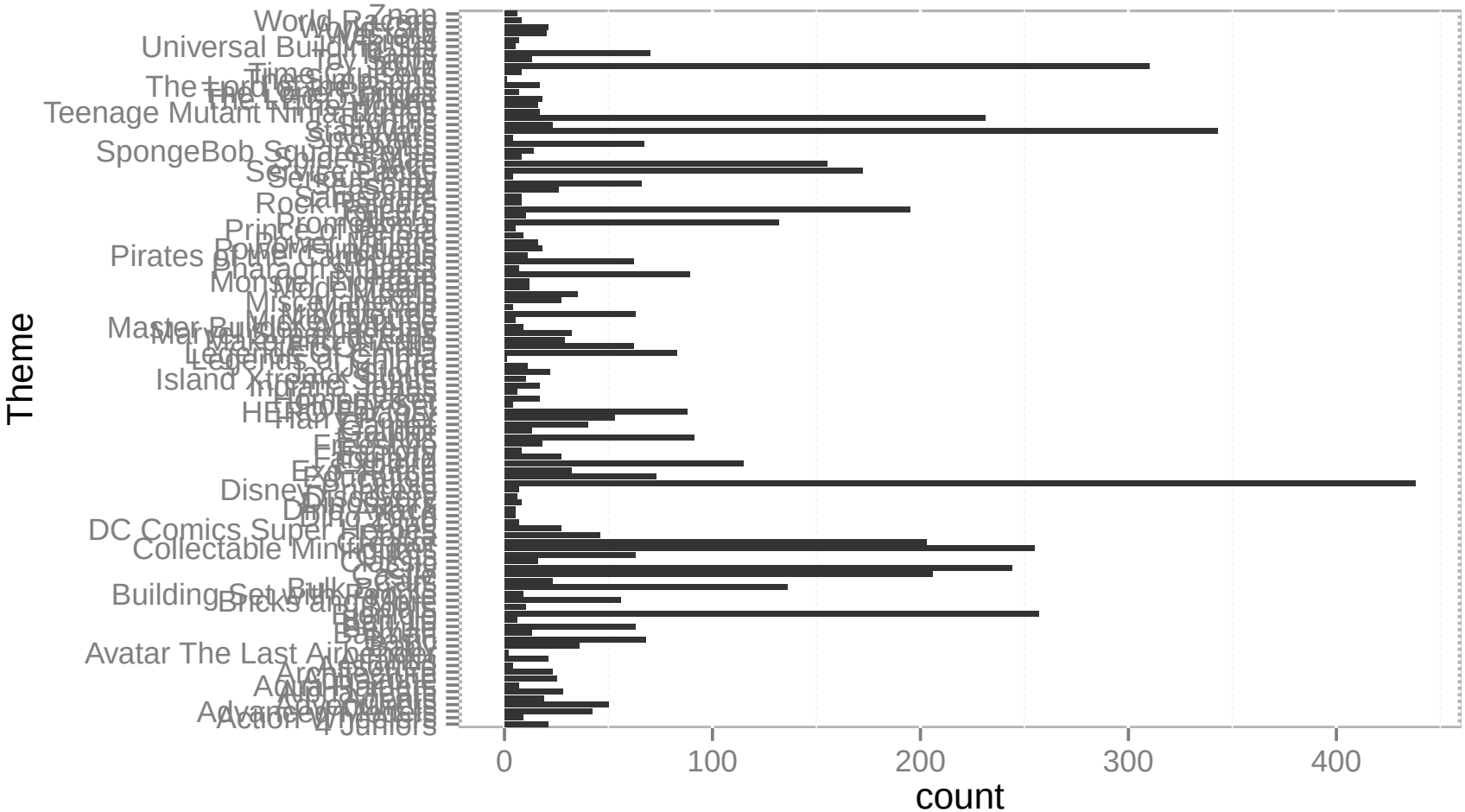
Barplots

```
ggplot(legosets, aes(x=Theme)) + geom_bar() +
    theme(axis.text.x=element_text(angle=90))
```



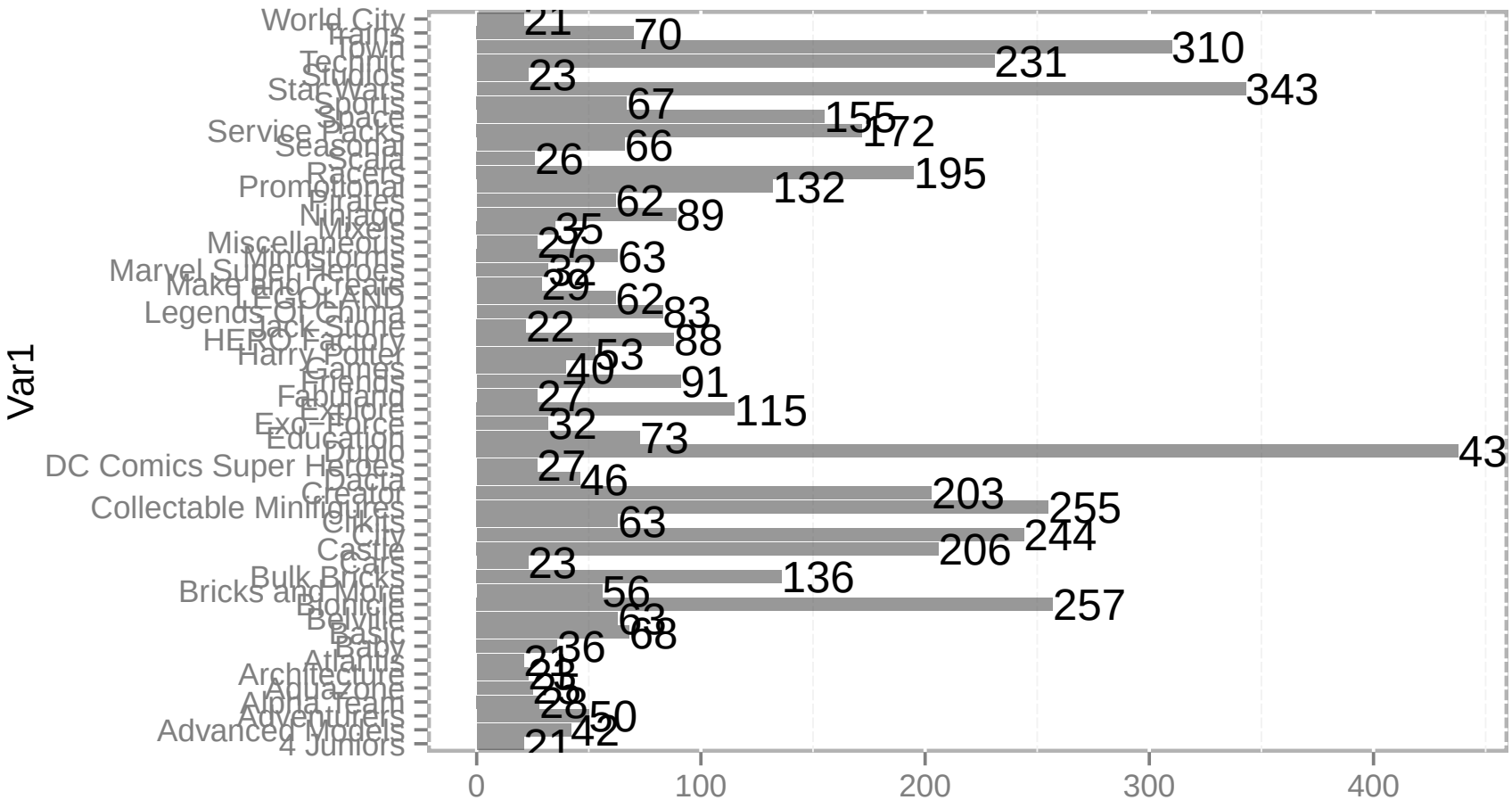
Barplots Flipping Coordinates

```
ggplot(legosets, aes(x=Theme)) + geom_bar() +
    coord_flip()
```



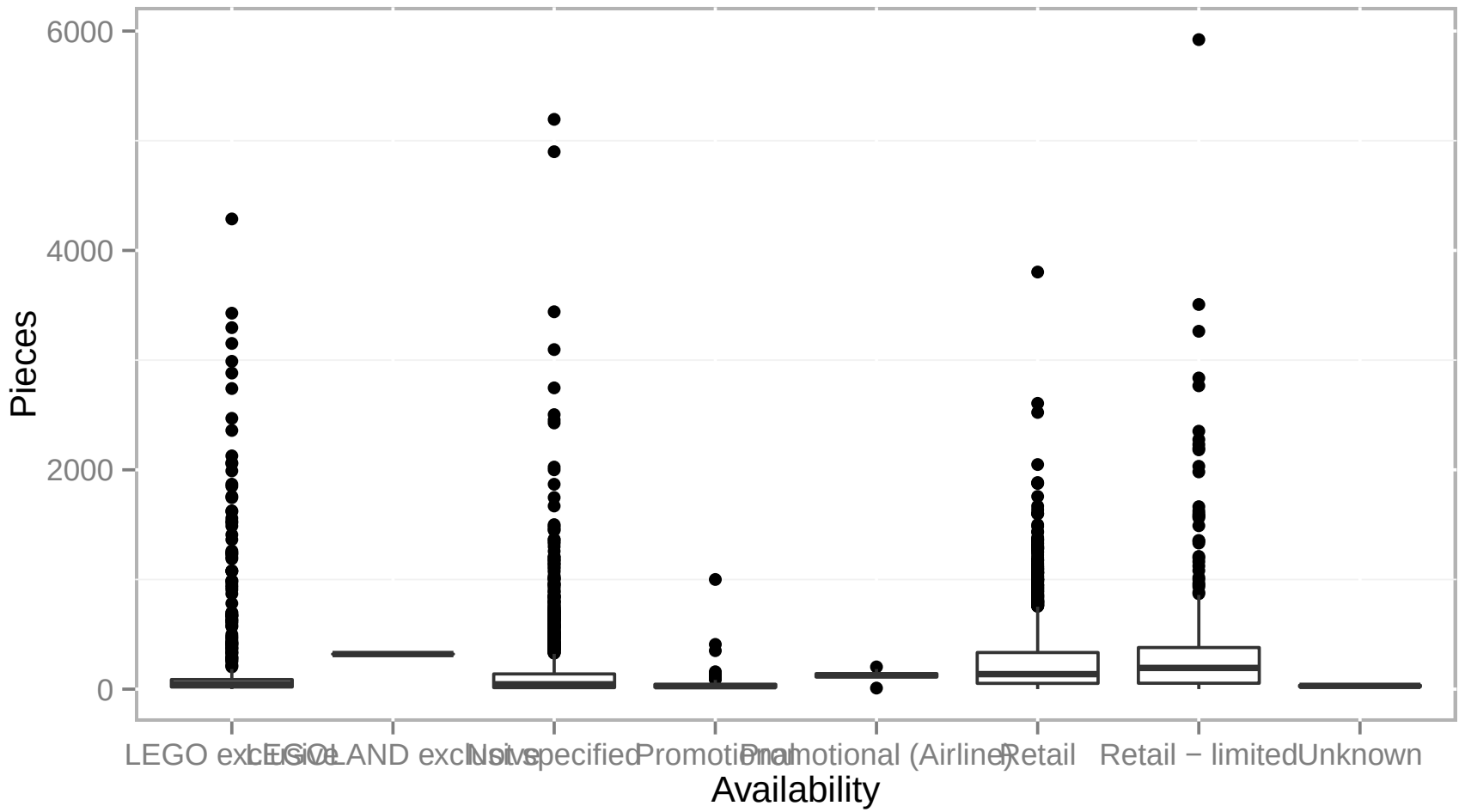
Labeling Barplots

```
df <- as.data.frame(table(legosets$Theme))
df <- df[df$Freq > 20,]
ggplot(df, aes(x=Var1, y=Freq, label=Freq)) +
    geom_bar(stat='identity', alpha=.5) +
    coord_flip() + geom_text(hjust=0)
```



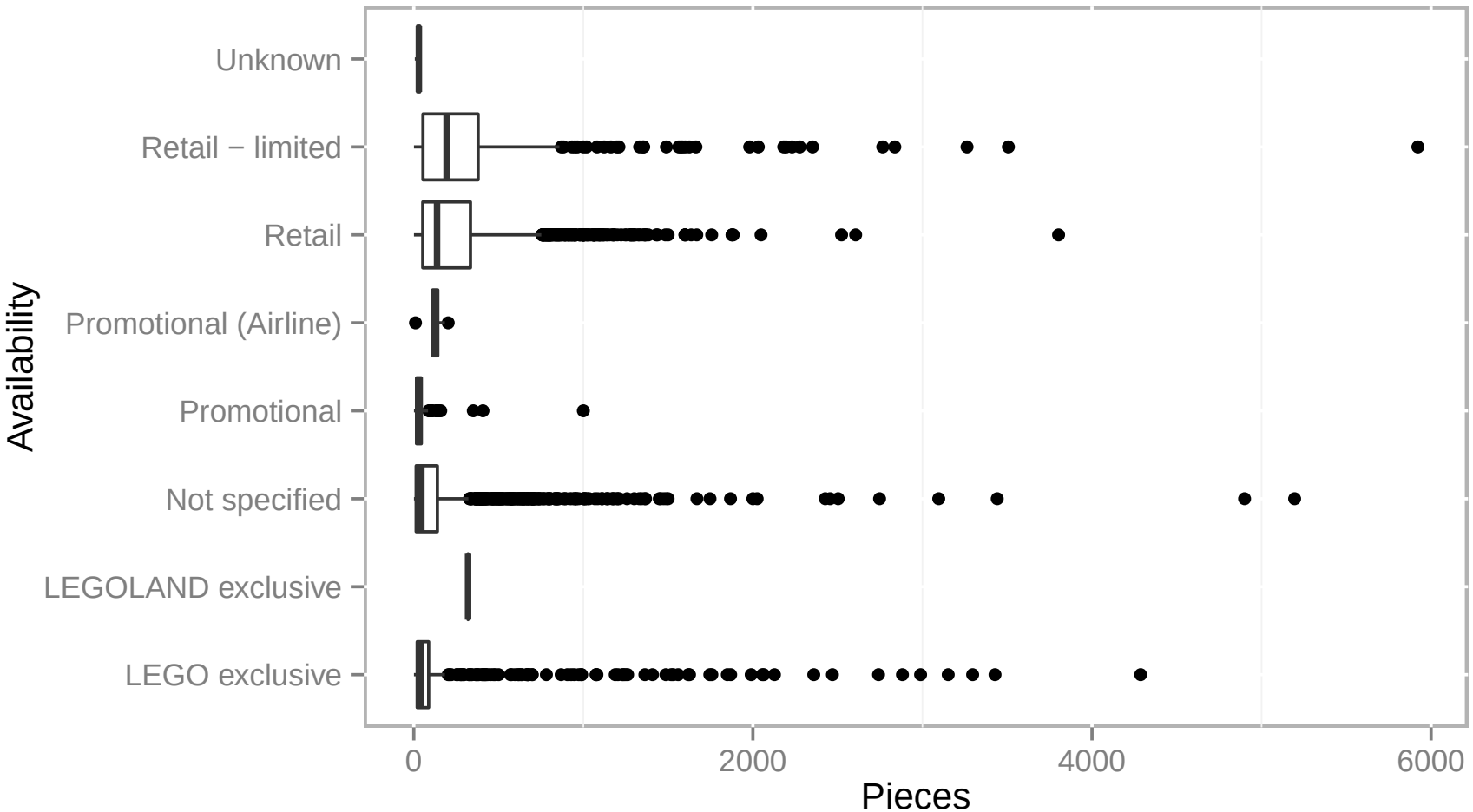
Boxplots

```
ggplot(legosets, aes(x=Availability, Pieces)) +
    geom_boxplot()
```



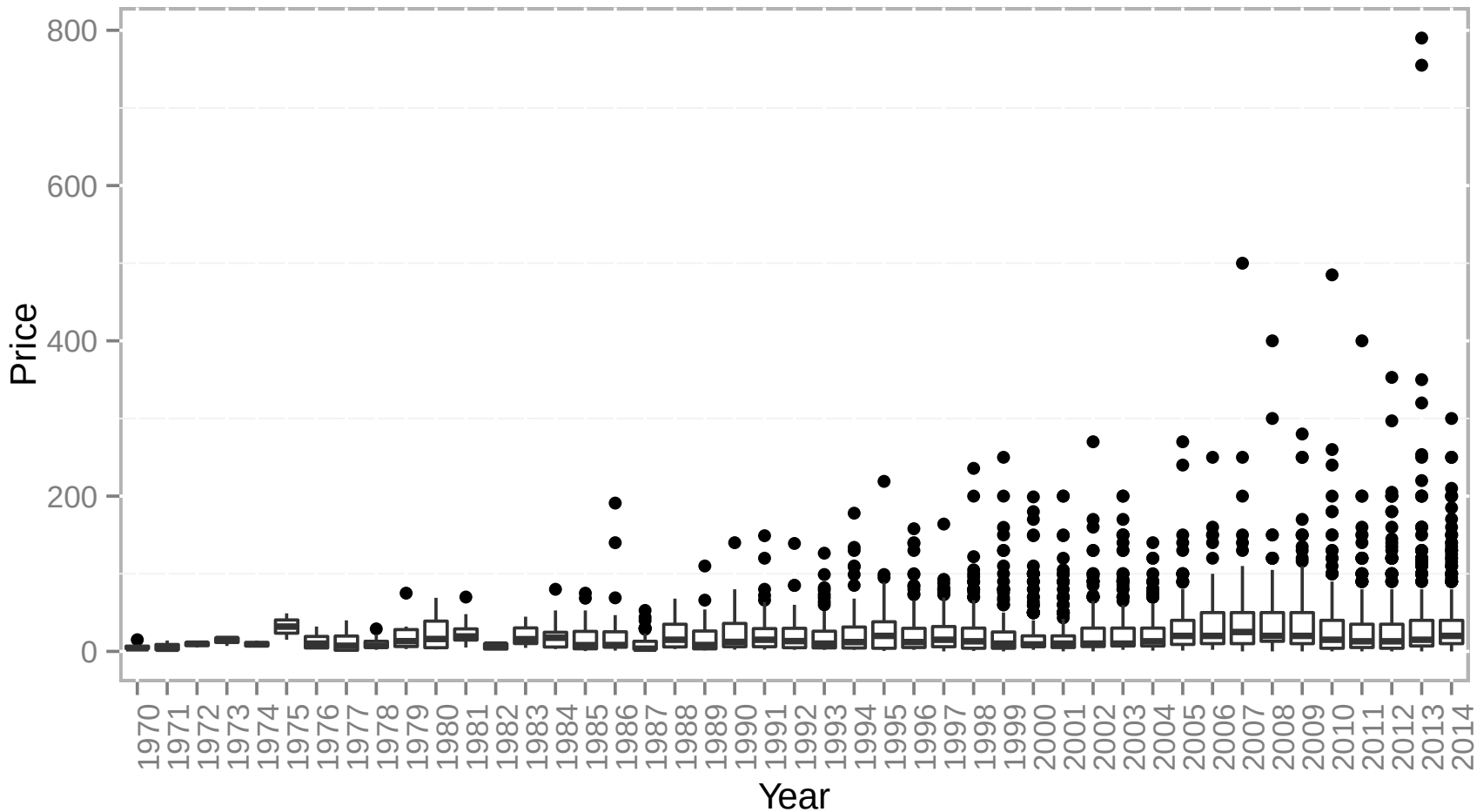
Boxplots Flipping Coordinates

```
ggplot(legosets, aes(x=Availability, Pieces)) +
    geom_boxplot() + coord_flip()
```



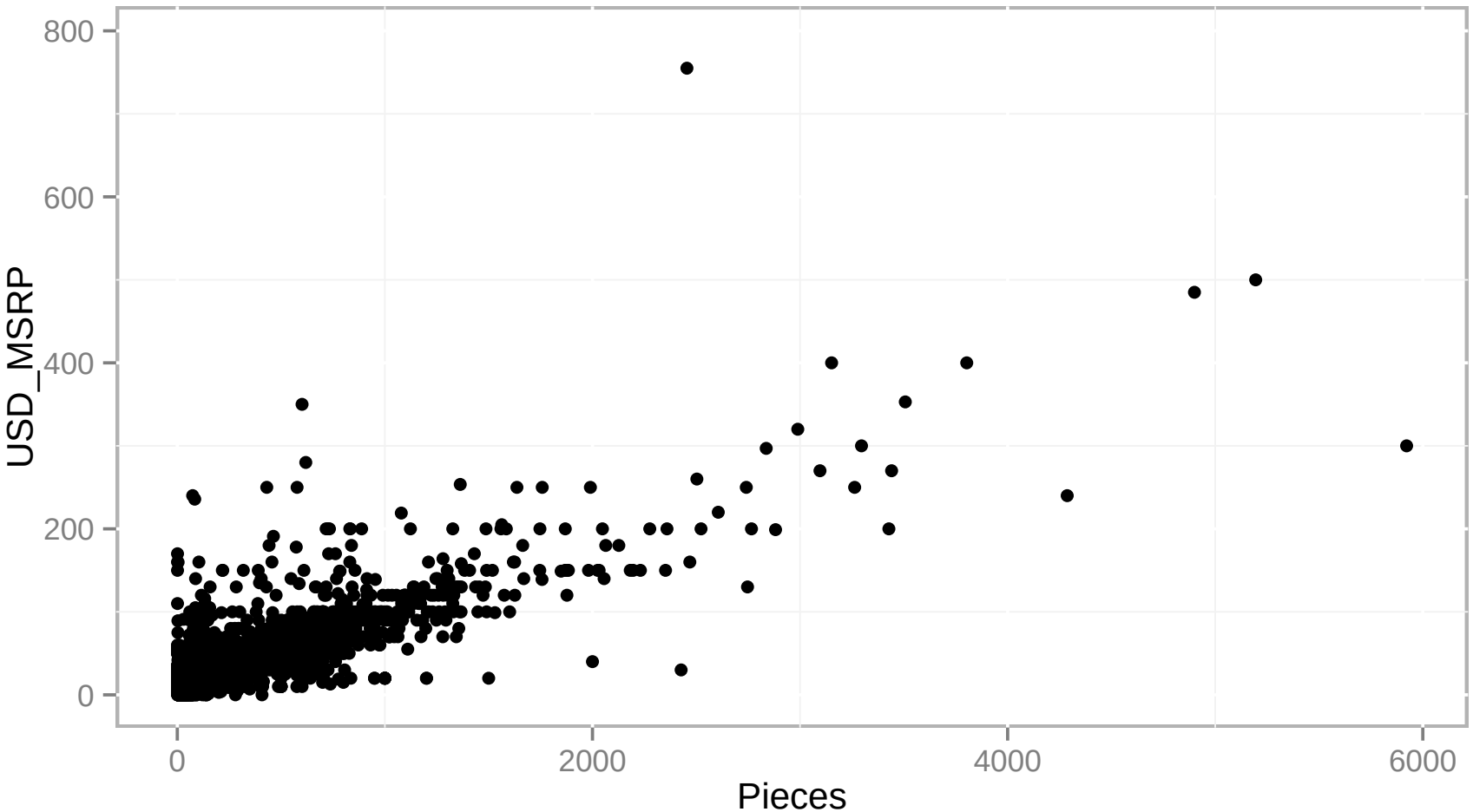
Boxplots for Longitudinal Data

```
ggplot(legosets, aes(x=factor(Year), y=USD_MSRP)) +
    geom_boxplot() + theme(axis.text.x=element_text(angle=9
    xlab('Year') + ylab('Price')
```



Scatterplots

```
ggplot(legosets, aes(x=Pieces, y=USD_MSRP)) +
    geom_point(alhpa=.5)
```



Scatterplots with Loess Plots

```
ggplot(legosets, aes(x=Pieces, y=USD_MSRP)) +
    geom_point(alhpa=.5) + geom_smooth()
```

```
## geom_smooth: method="auto" and size of largest group is
```