

Exploration of Prosper Loan Data by Patrick Spears

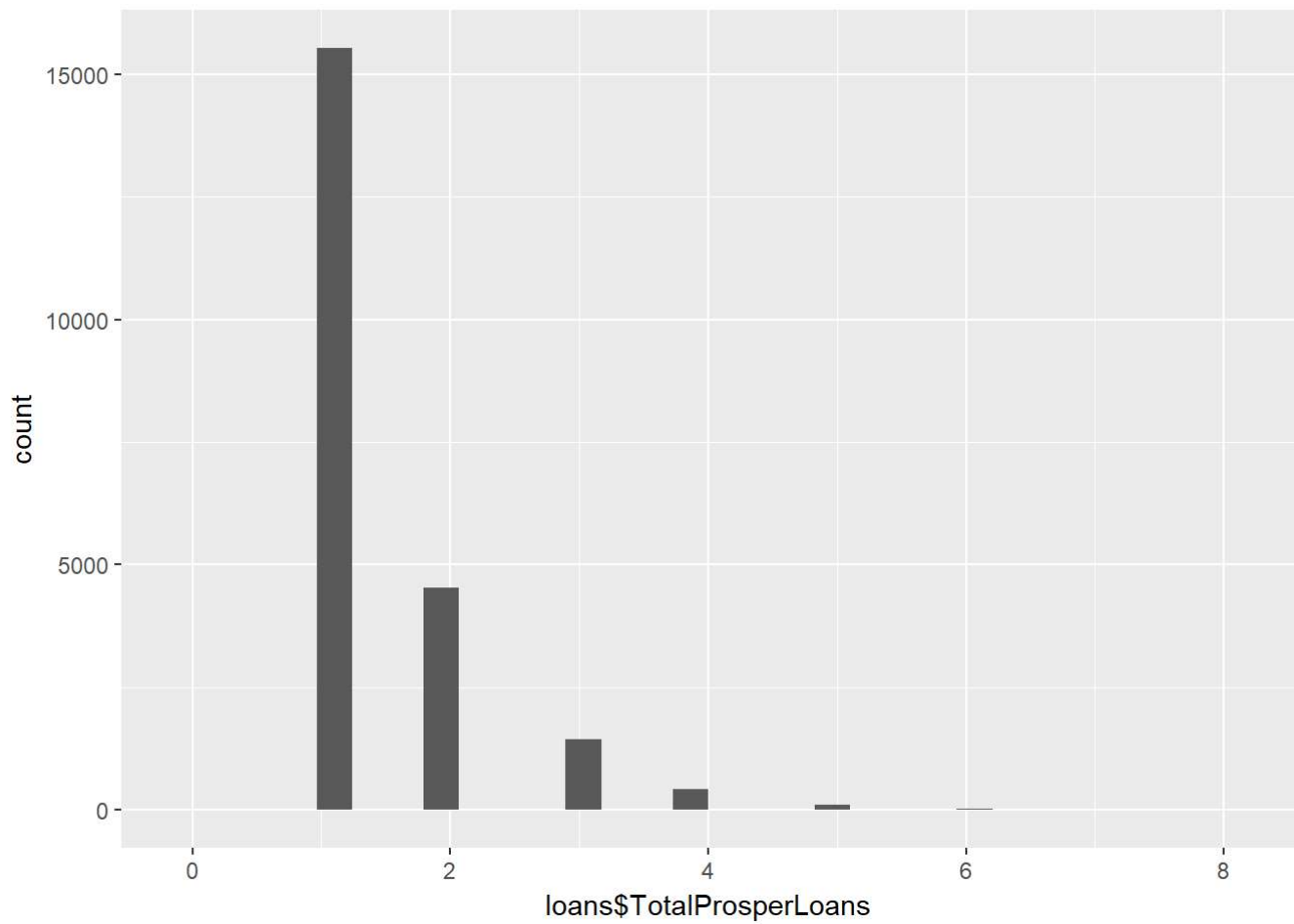
I will be exploring the Prosper loan data set, a collection of loans made through the fintech company Prosper. Prosper describes itself as a peer-to-peer lending platform (or “marketplace”): according to the “About Us” section of the website, “individuals and institutions can invest in loans and earn attractive returns.” The data includes the details of 113,937 loans, each with 81 attributes.

Univariate Plots Section

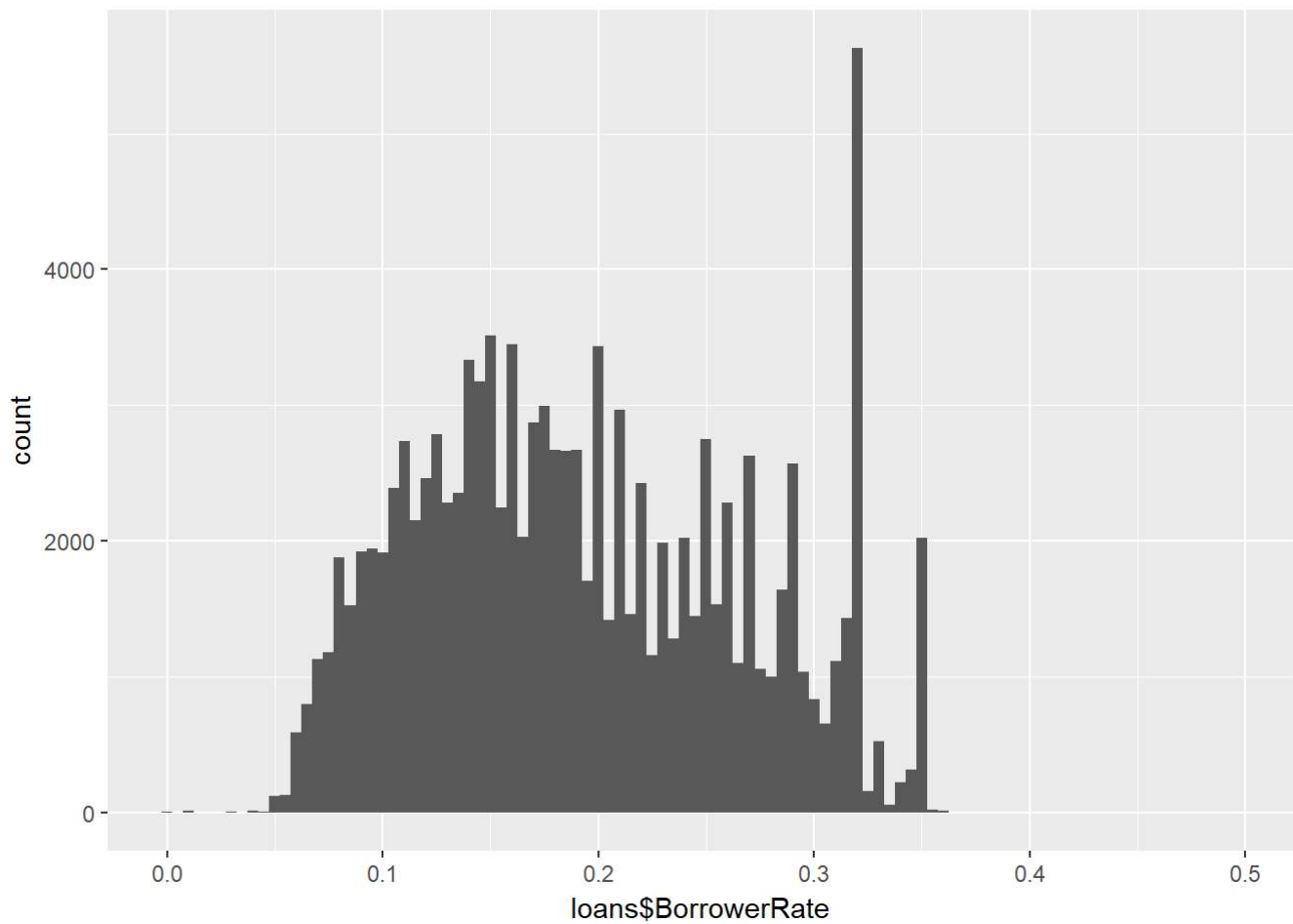
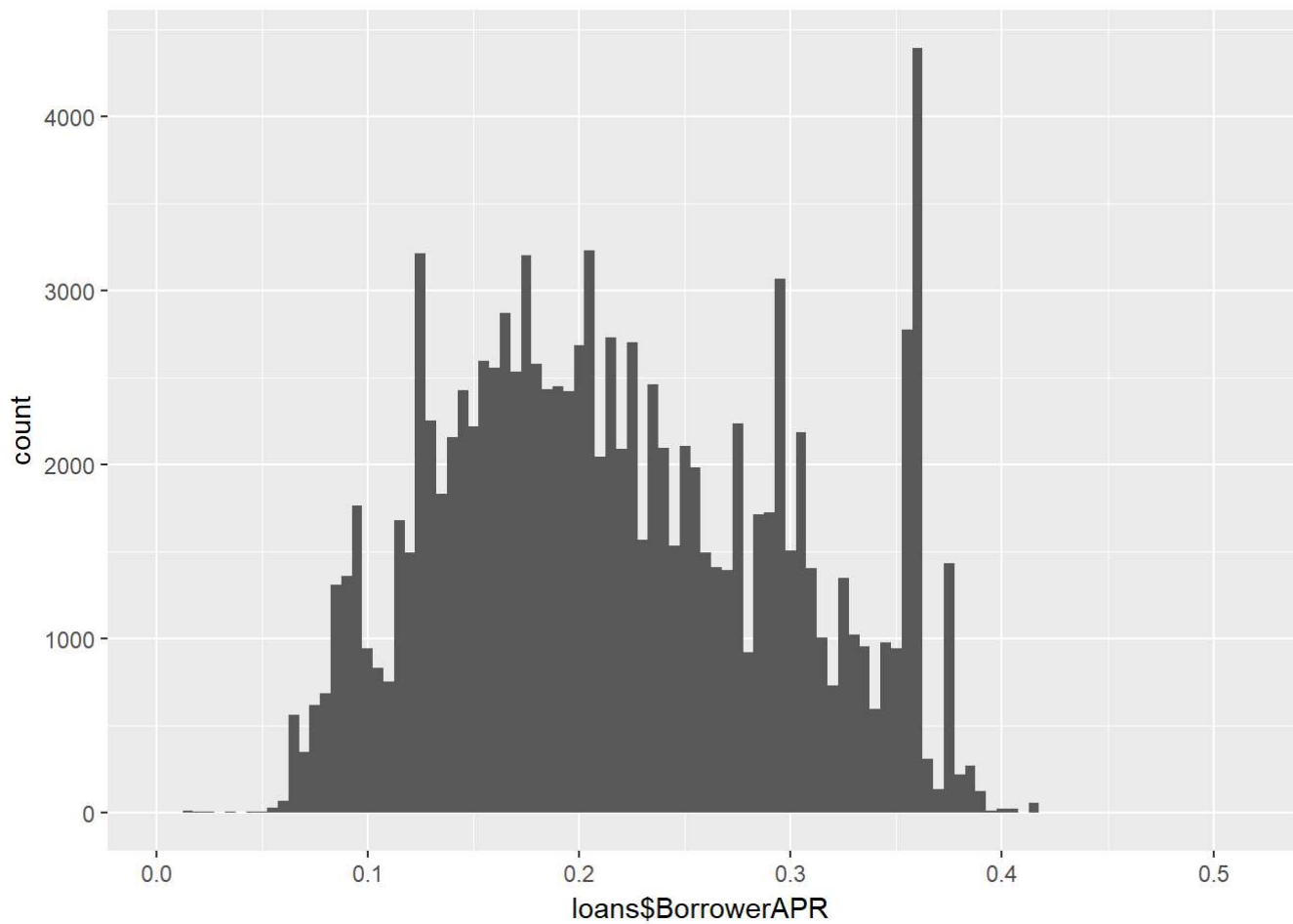
Tip: In this section, you should perform some preliminary exploration of your dataset. Run some summaries of the data and create univariate plots to understand the structure of the individual variables in your dataset. Don’t forget to add a comment after each plot or closely-related group of plots! There should be multiple code chunks and text sections; the first one below is just to help you get started.

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

```
## Warning: Removed 91852 rows containing non-finite values (stat_bin).
```

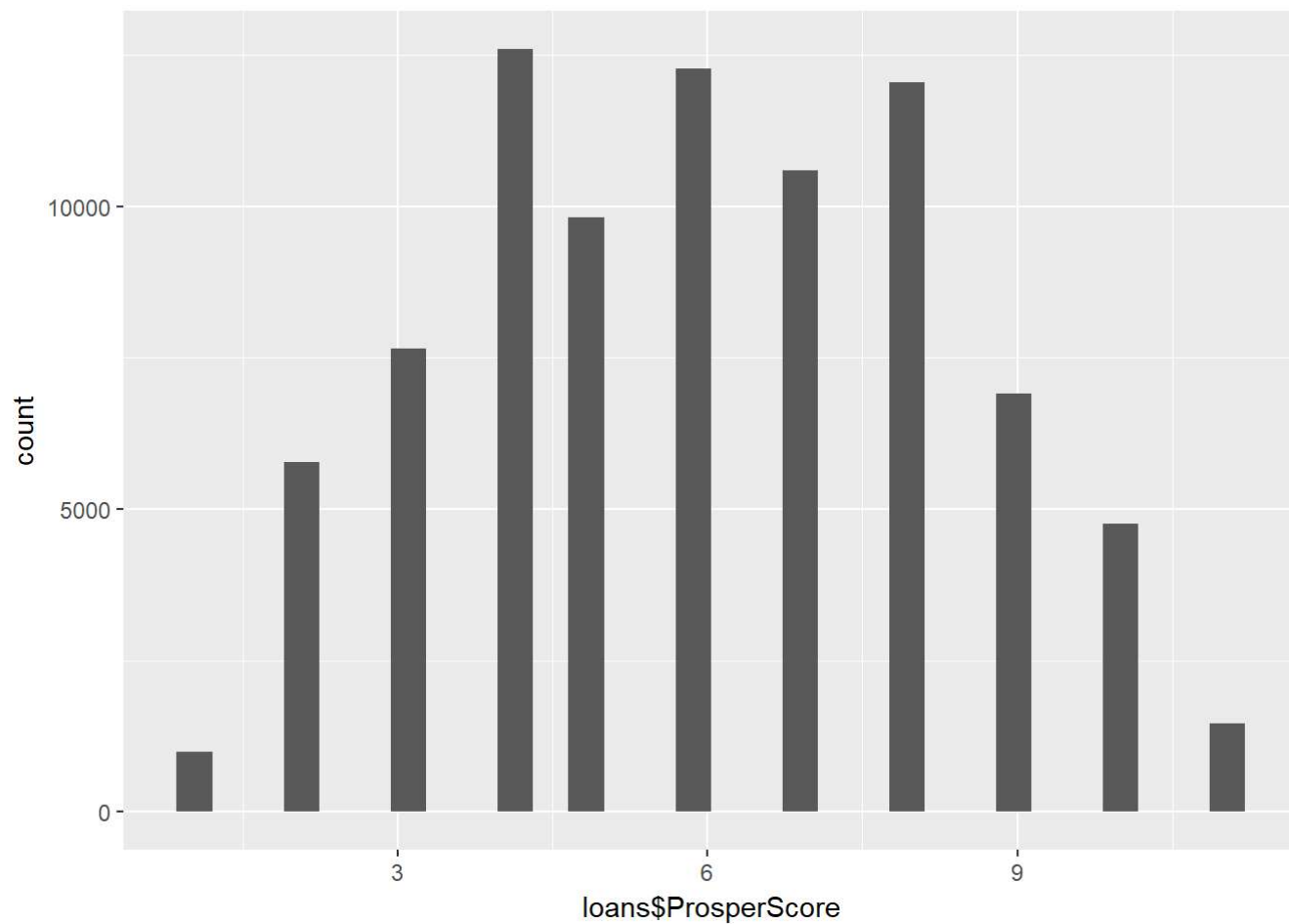


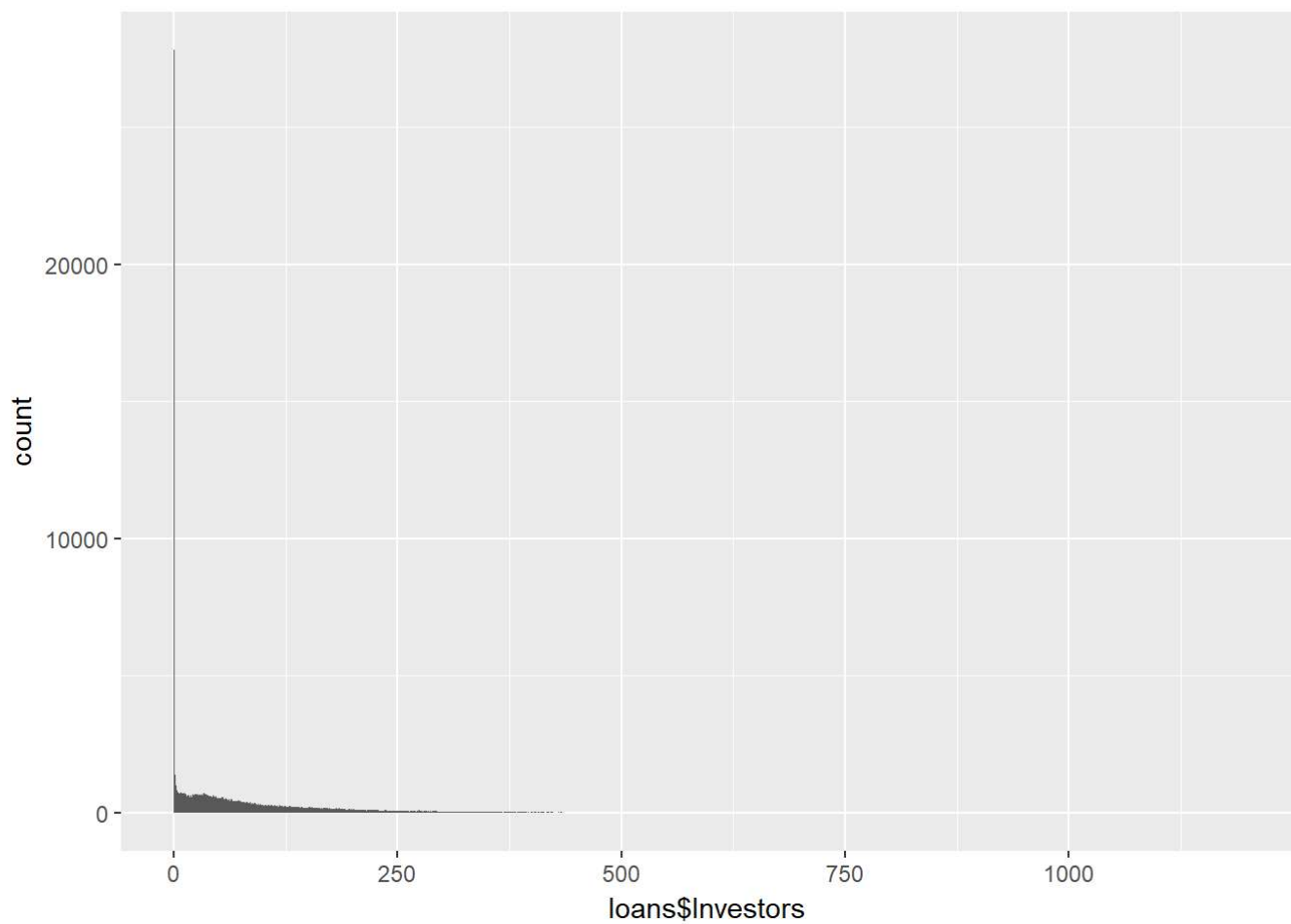
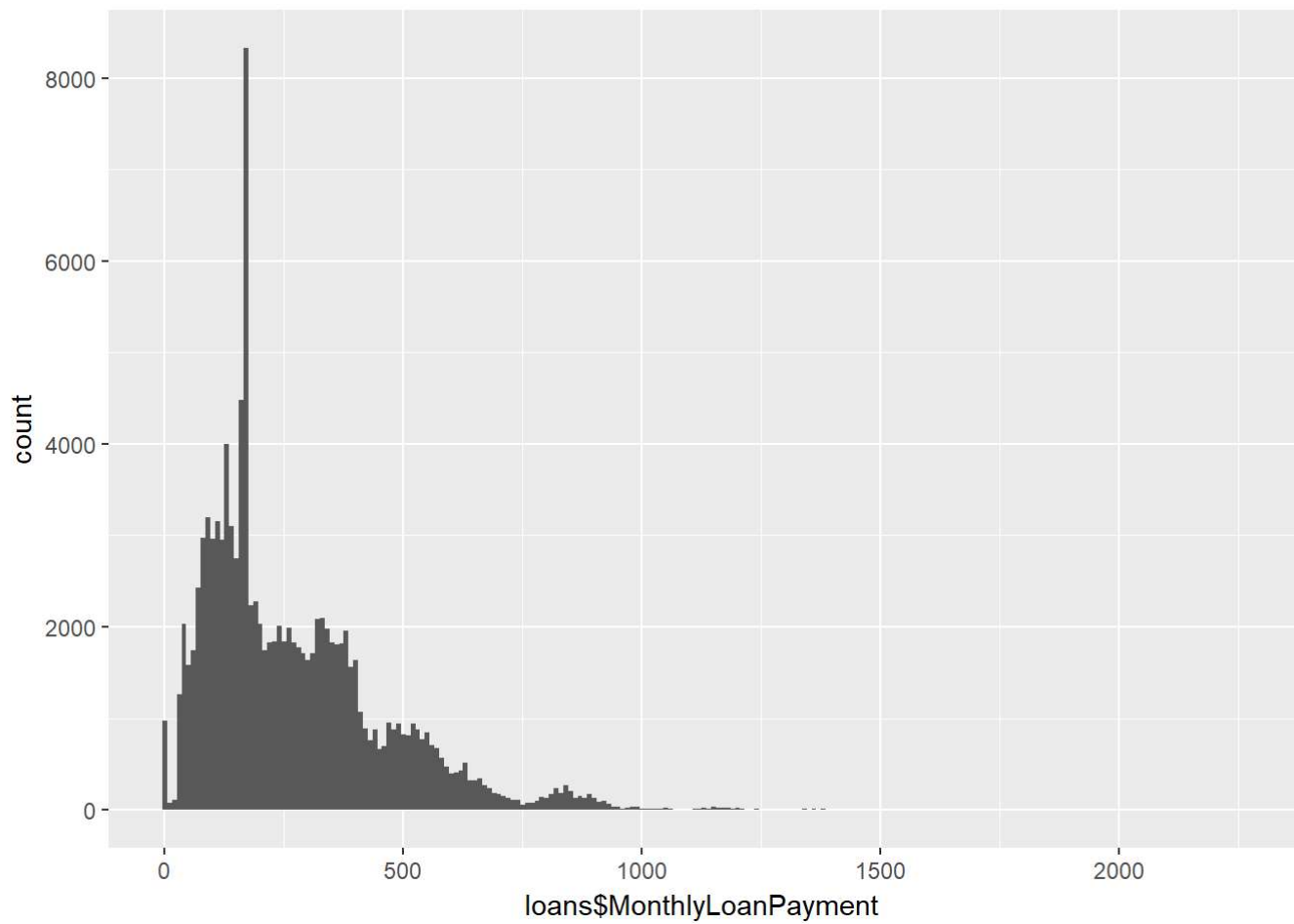
```
## Warning: Removed 25 rows containing non-finite values (stat_bin).
```



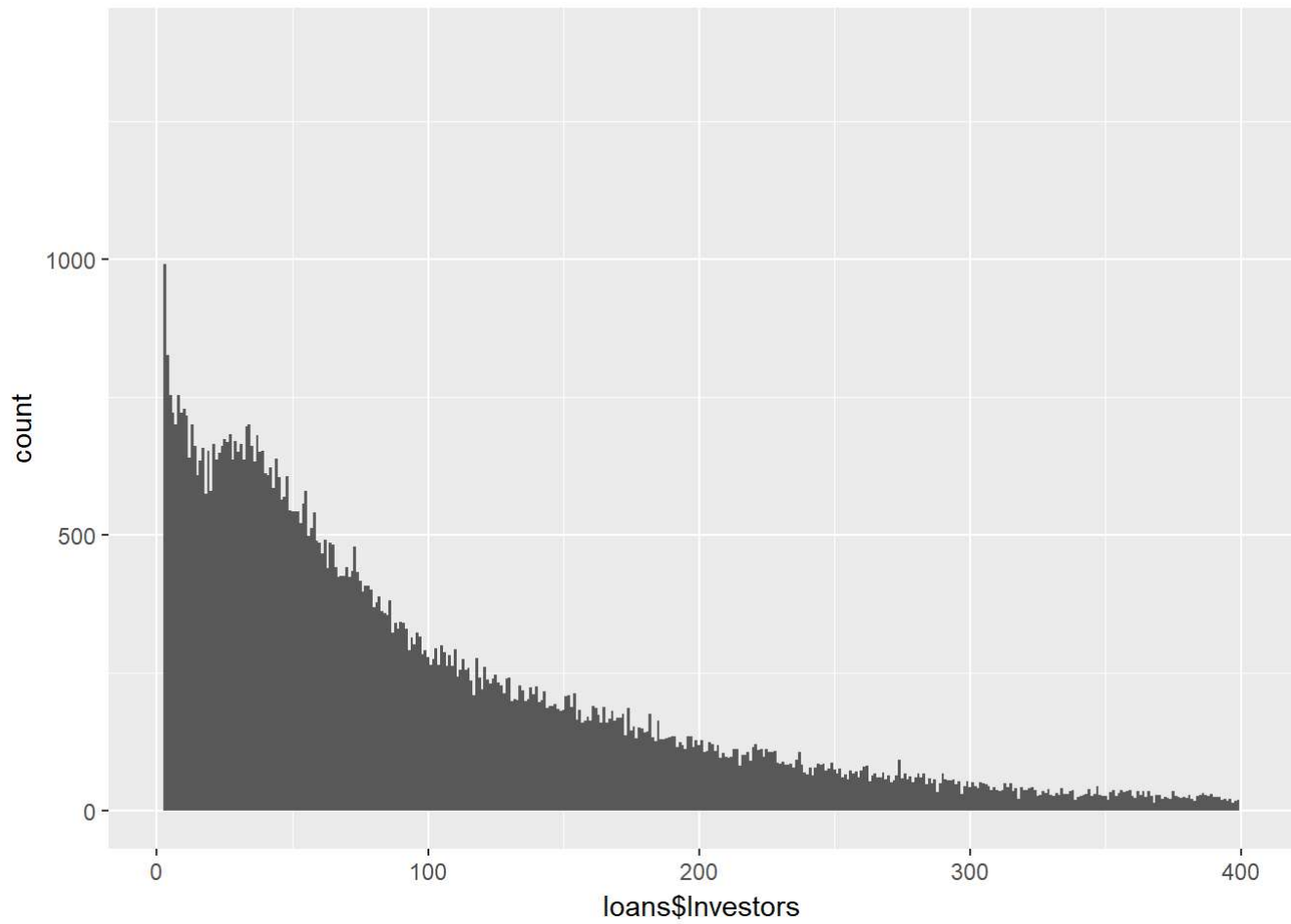
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

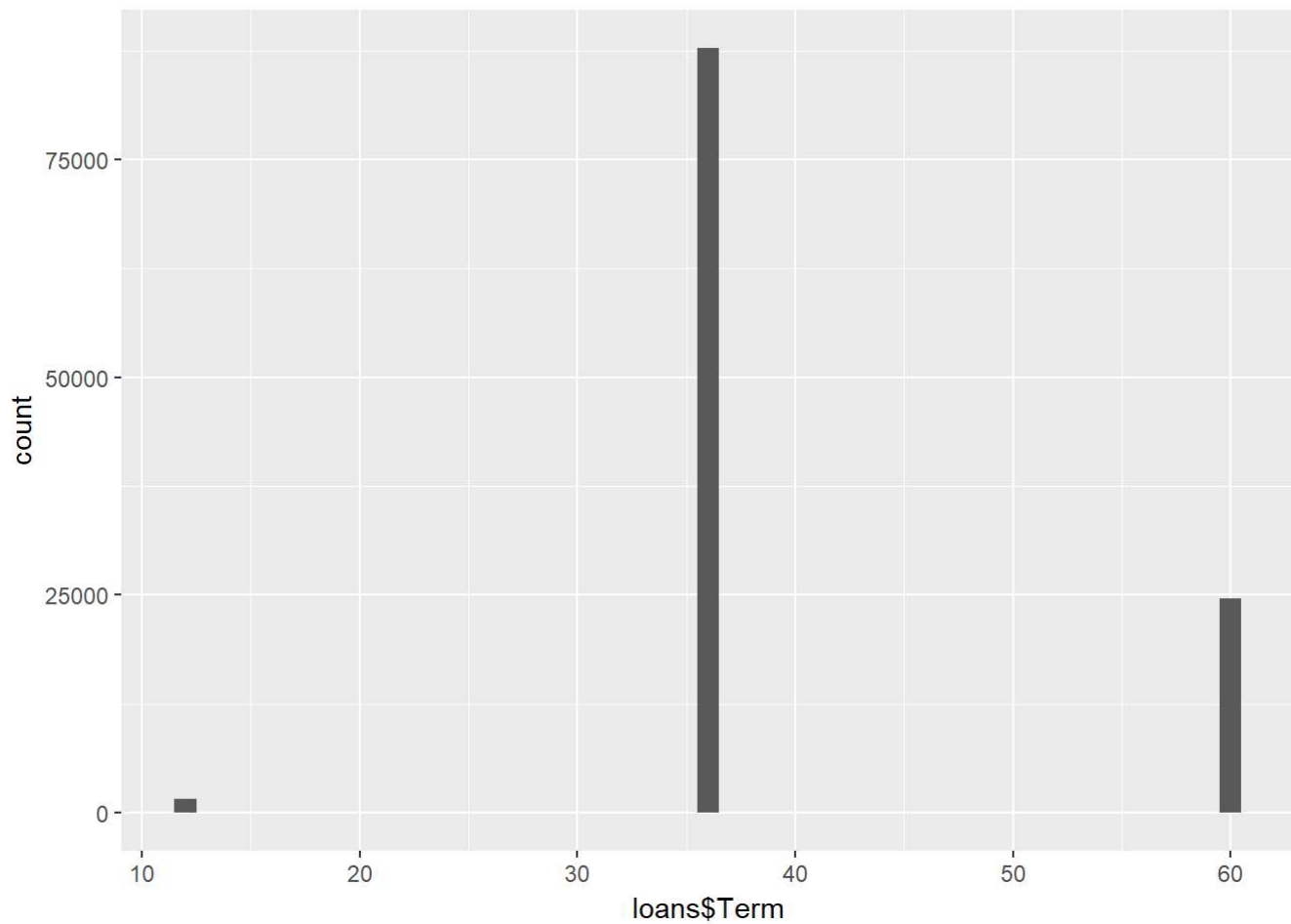
```
## Warning: Removed 29084 rows containing non-finite values (stat_bin).
```





```
## Warning: Removed 30000 rows containing non-finite values (stat_bin).
```

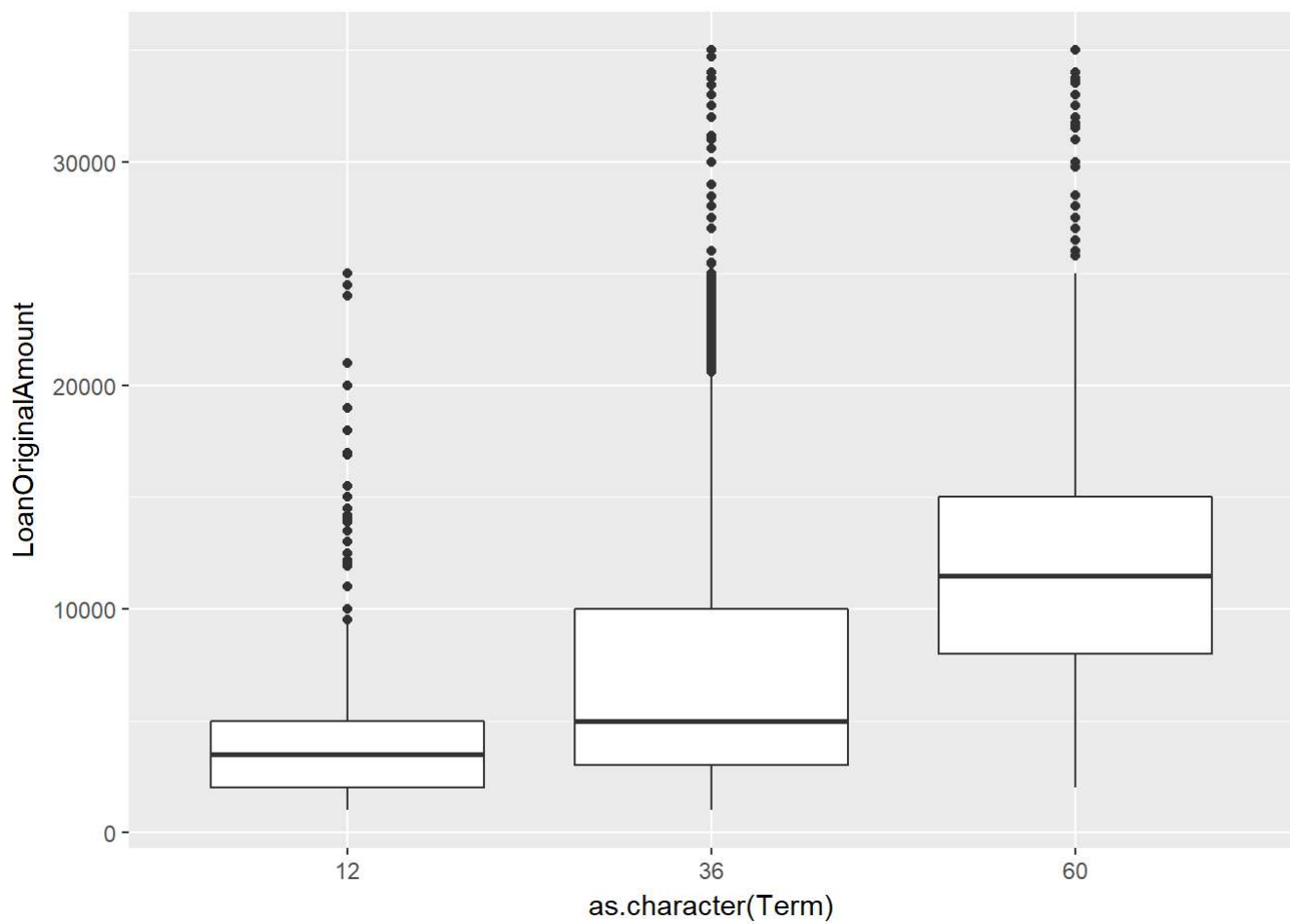
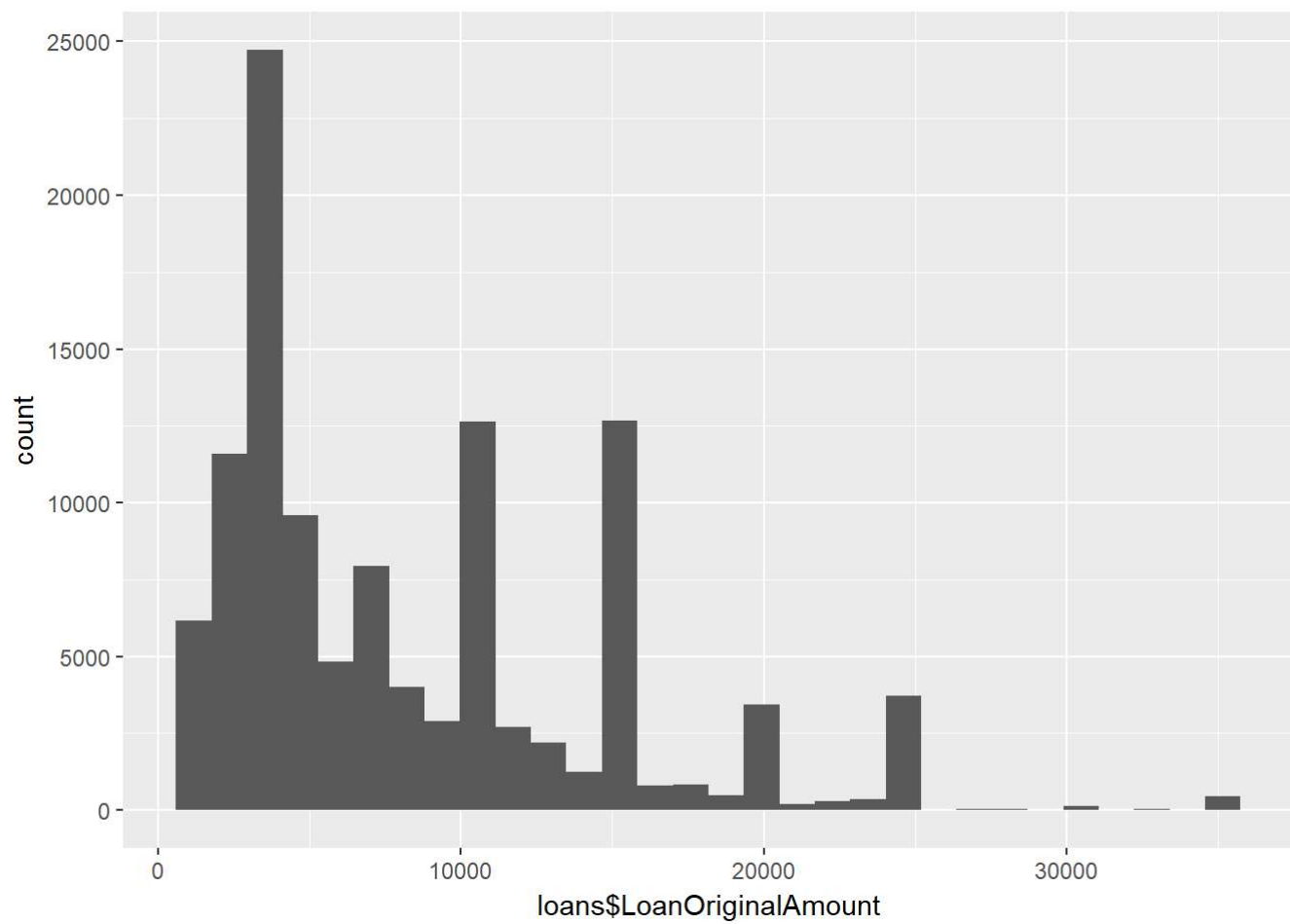


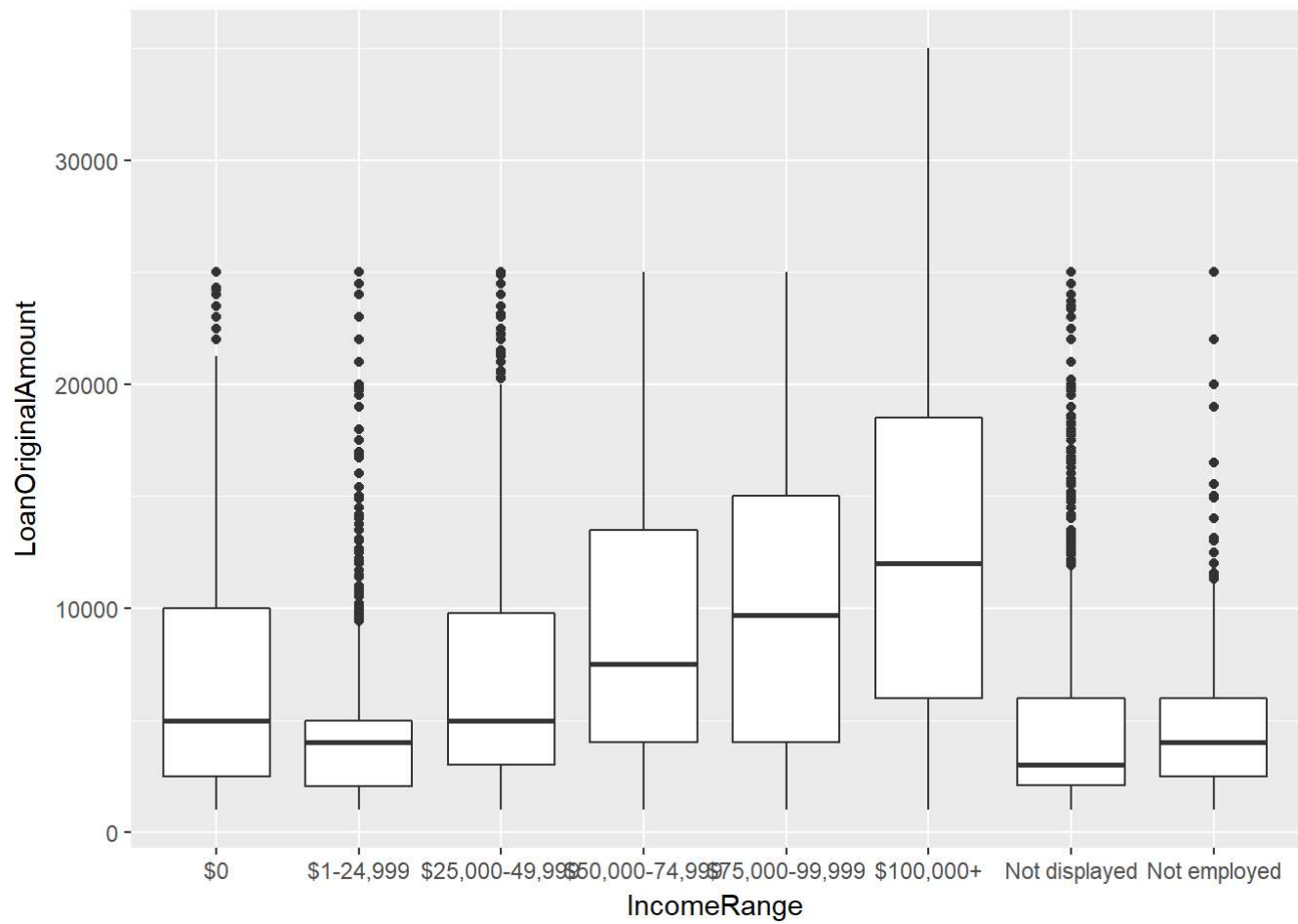


```
## [1] 36 60 12
```

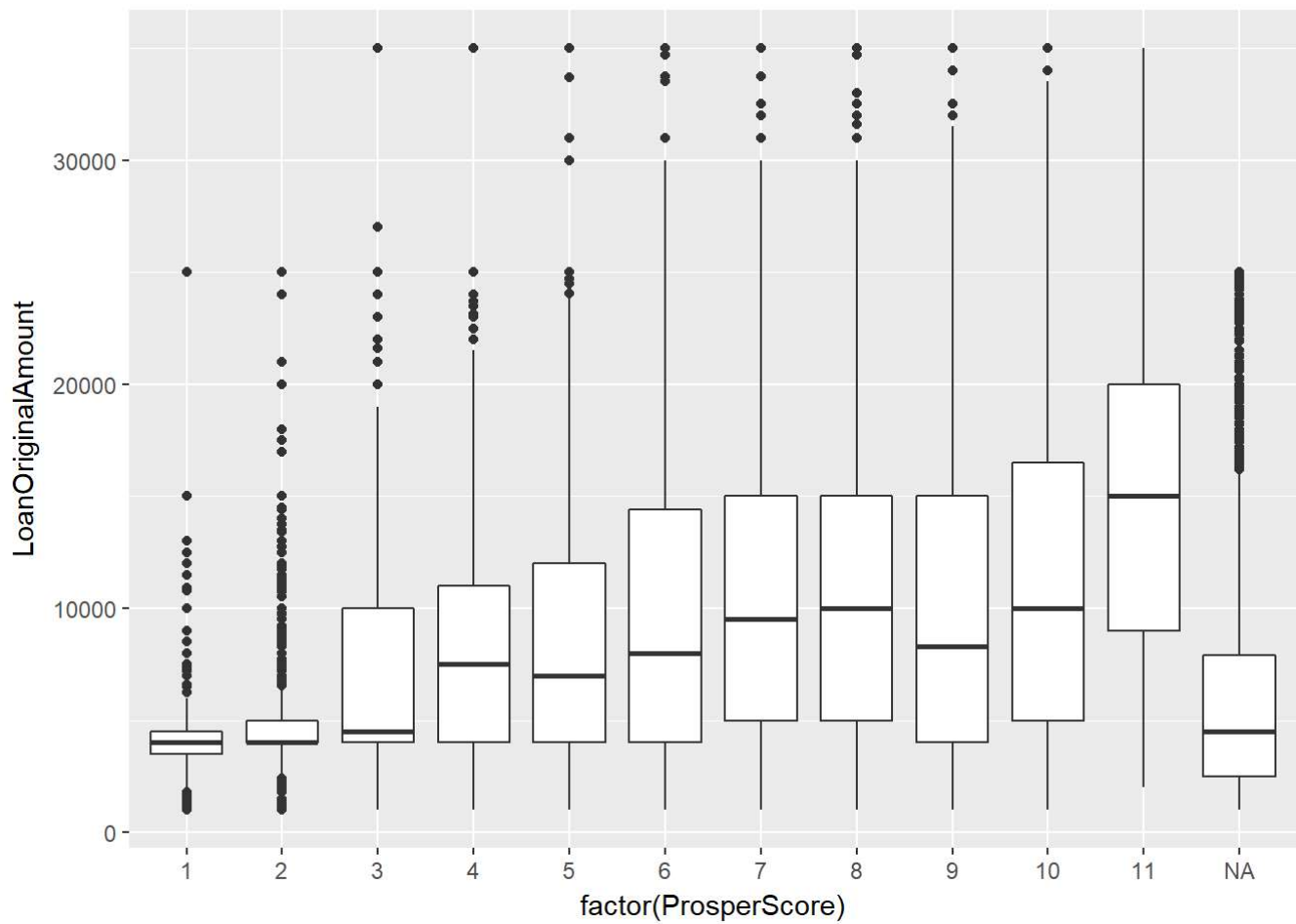
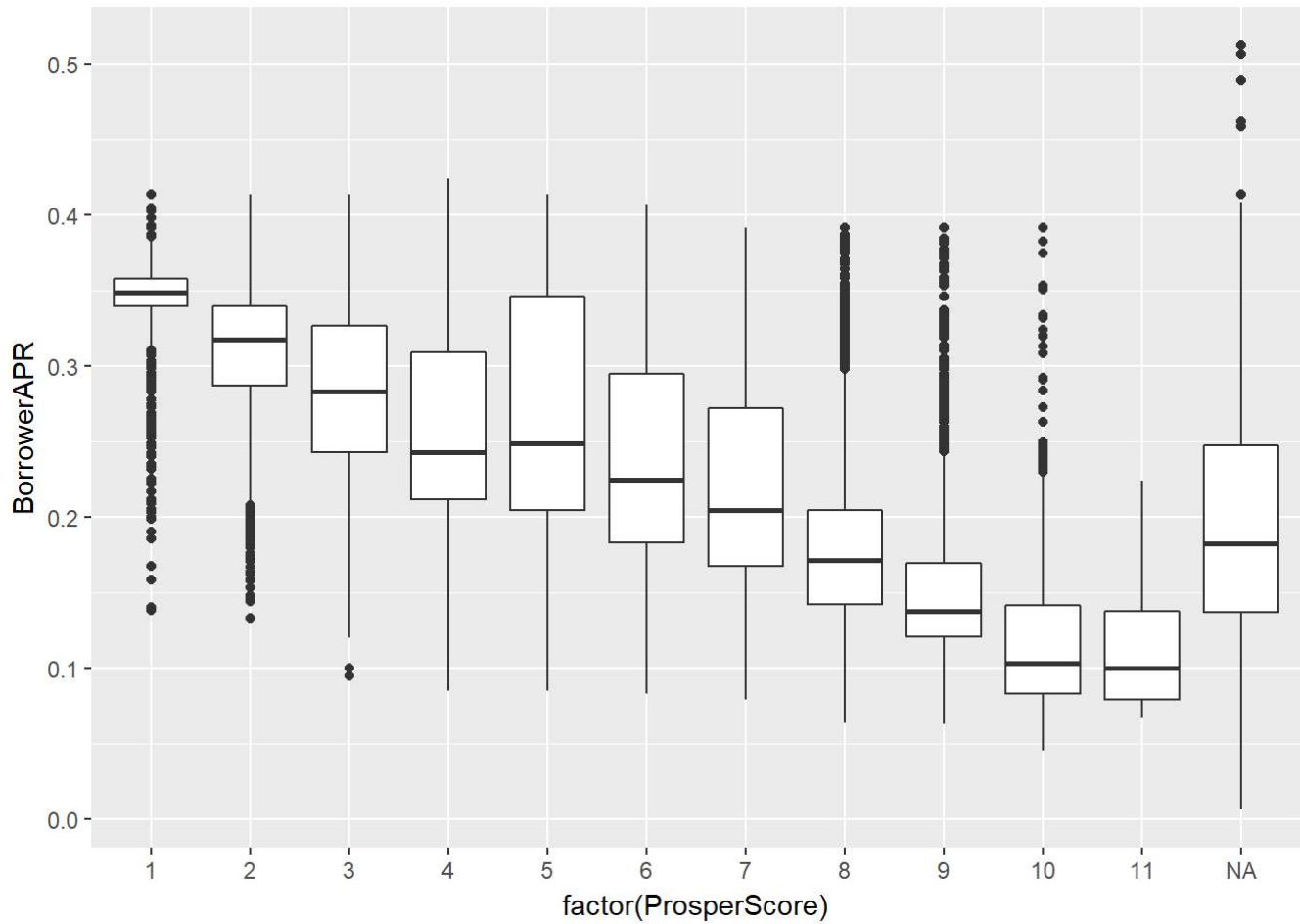
```
## Warning: Ignoring unknown parameters: binwidth
```

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

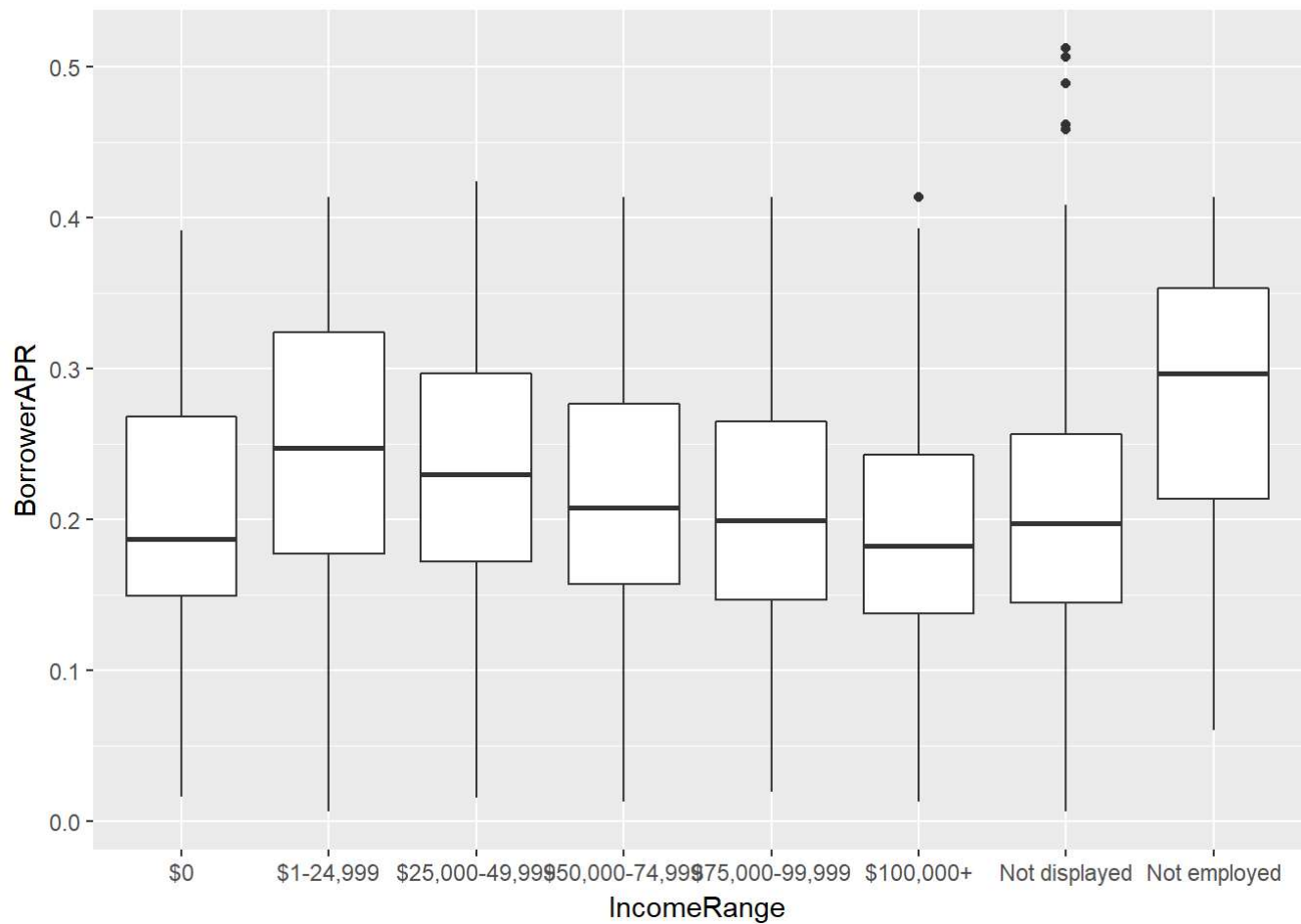




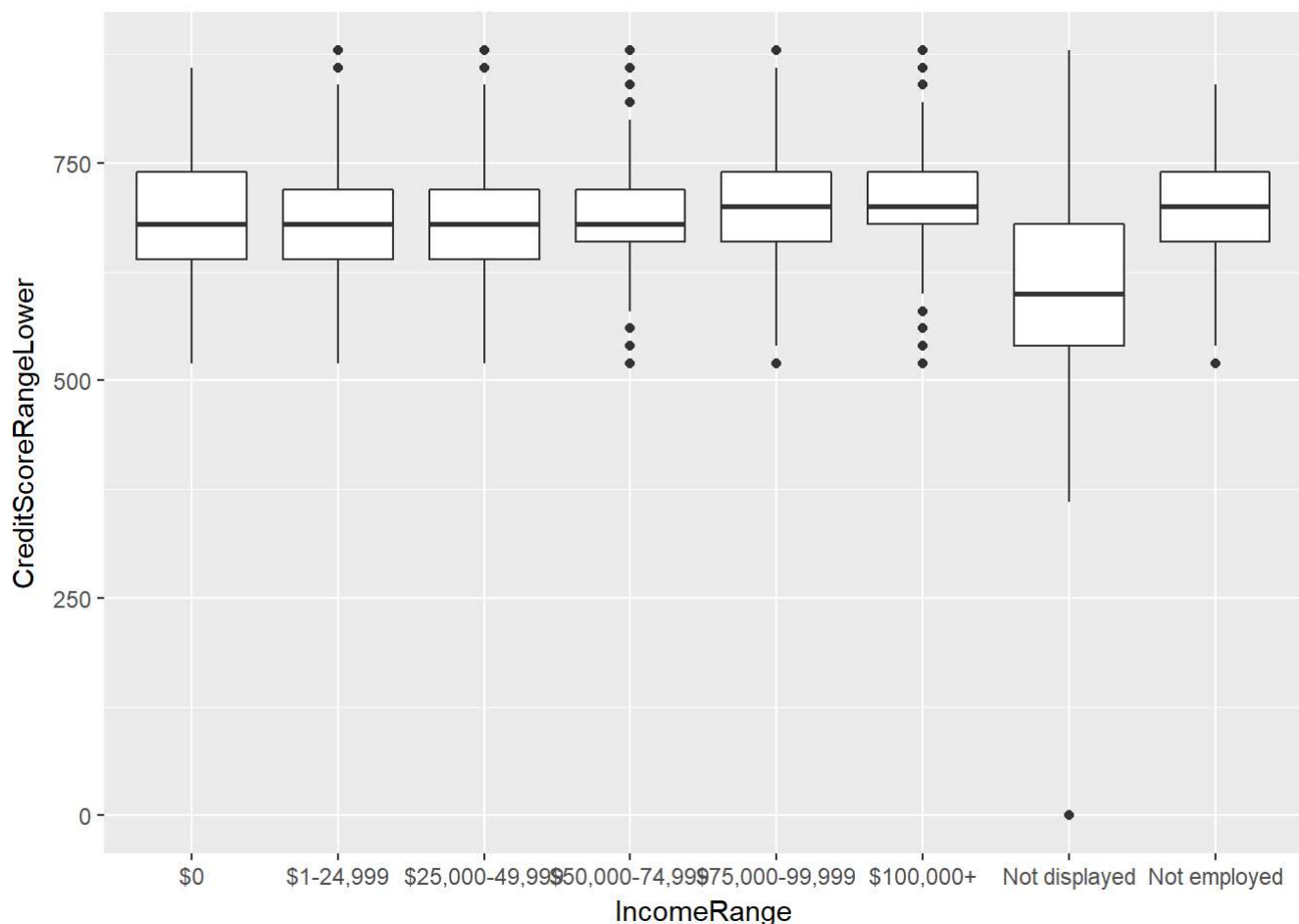
```
## Warning: Removed 25 rows containing non-finite values (stat_boxplot).
```



```
## Warning: Removed 25 rows containing non-finite values (stat_boxplot).
```



```
## Warning: Removed 591 rows containing non-finite values (stat_boxplot).
```



```
## [1] 640 680 480 800 740 700 820 760 660 620 720 520 780 600 580 540 560
## [18] 500 840 860 NA 460 0 880 440 420 360
```

Tip: Make sure that you leave a blank line between the start / end of each code block and the end / start of your Markdown text so that it is formatted nicely in the knitted text. Note as well that text on consecutive lines is treated as a single space. Make sure you have a blank line between your paragraphs so that they too are formatted for easy readability.

Univariate Analysis

Tip: Now that you've completed your univariate explorations, it's time to reflect on and summarize what you've found. Use the questions below to help you gather your observations and add your own if you have other thoughts!

What is the structure of your dataset?

What is/are the main feature(s) of interest in your dataset?

What other features in the dataset do you think will help support your investigation into your feature(s) of interest?

Did you create any new variables from existing variables in the dataset?

Of the features you investigated, were there any unusual distributions?

Did you perform any operations on the data to tidy, adjust, or change the form of the data? If so, why did you do this?

Bivariate Plots Section

Tip: Based on what you saw in the univariate plots, what relationships between variables might be interesting to look at in this section? Don't limit yourself to relationships between a main output feature and one of the supporting variables. Try to look at relationships between supporting variables as well.

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 25 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 591 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 8554 rows containing missing values
```

```
## Warning: Removed 25 rows containing missing values (geom_point).
```

```
## Warning: Removed 25 rows containing non-finite values (stat_density).
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 25 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 25 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 25 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 591 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 8579 rows containing missing values
```

```
## Warning: Removed 25 rows containing missing values (geom_point).
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 591 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 8554 rows containing missing values
```

```
## Warning: Removed 25 rows containing missing values (geom_point).
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 591 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 8554 rows containing missing values
```

```
## Warning: Removed 25 rows containing missing values (geom_point).
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 591 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 8554 rows containing missing values
```

```
## Warning: Removed 591 rows containing missing values (geom_point).
```

```
## Warning: Removed 591 rows containing missing values (geom_point).
```

```
## Warning: Removed 591 rows containing missing values (geom_point).
```

```
## Warning: Removed 591 rows containing missing values (geom_point).
```

```
## Warning: Removed 591 rows containing missing values (geom_point).
```

```
## Warning: Removed 591 rows containing non-finite values (stat_density).
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 9139 rows containing missing values
```

```
## Warning: Removed 8554 rows containing missing values (geom_point).
```

```
## Warning: Removed 8579 rows containing missing values (geom_point).
```

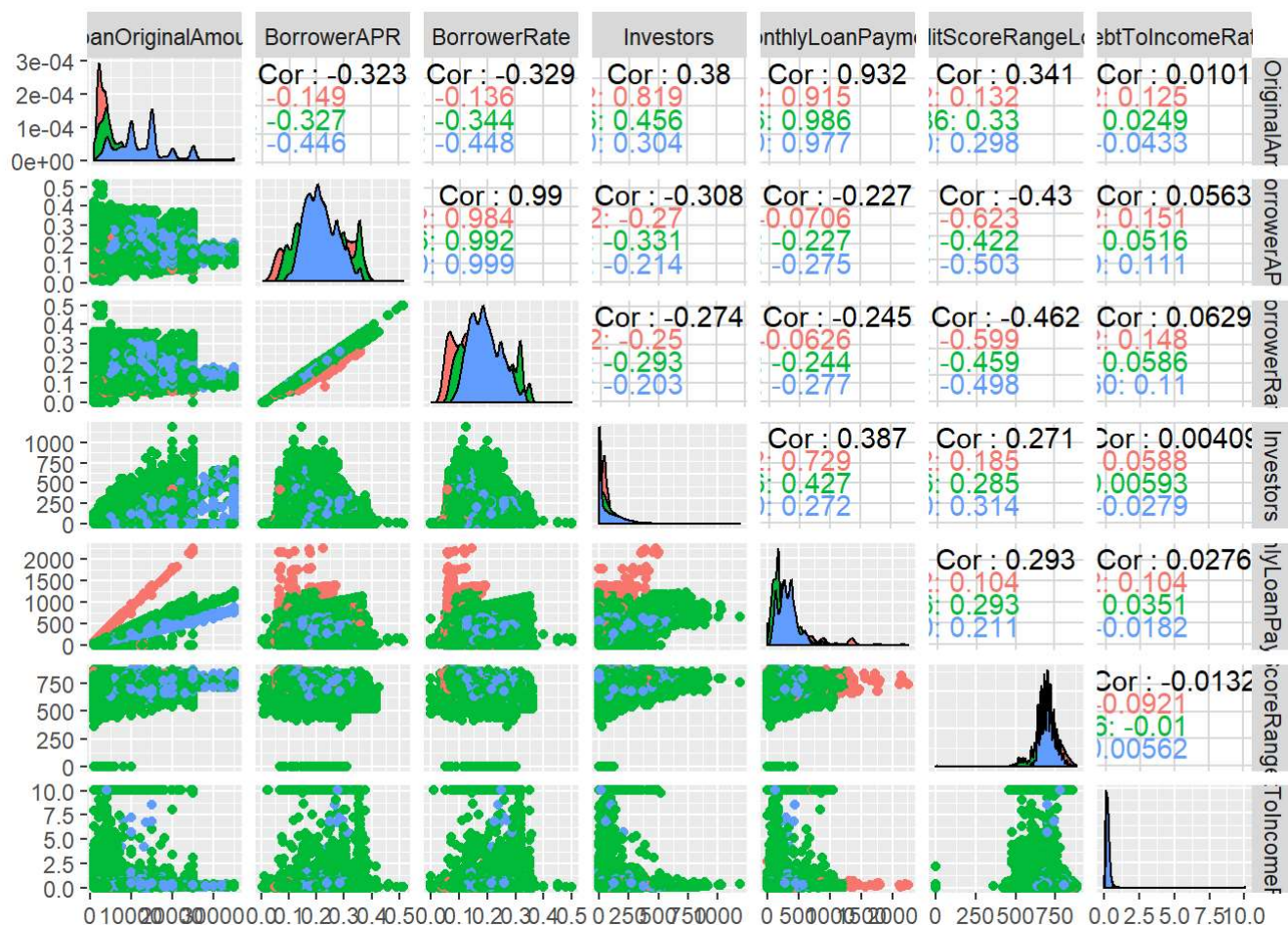
```
## Warning: Removed 8554 rows containing missing values (geom_point).
```

```
## Warning: Removed 8554 rows containing missing values (geom_point).
```

```
## Warning: Removed 8554 rows containing missing values (geom_point).
```

```
## Warning: Removed 9139 rows containing missing values (geom_point).
```

```
## Warning: Removed 8554 rows containing non-finite values (stat_density).
```



```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 25 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 591 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 8554 rows containing missing values
```

```
## Warning: Removed 25 rows containing missing values (geom_point).
```

```
## Warning: Removed 25 rows containing non-finite values (stat_density).
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 25 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 25 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 25 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 591 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 8579 rows containing missing values
```

```
## Warning: Removed 25 rows containing missing values (geom_point).
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 591 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 8554 rows containing missing values
```

```
## Warning: Removed 25 rows containing missing values (geom_point).
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 591 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 8554 rows containing missing values
```

```
## Warning: Removed 25 rows containing missing values (geom_point).
```



```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 591 rows containing missing values
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 8554 rows containing missing values
```

```
## Warning: Removed 591 rows containing missing values (geom_point).  
  
## Warning: Removed 591 rows containing missing values (geom_point).  
  
## Warning: Removed 591 rows containing missing values (geom_point).  
  
## Warning: Removed 591 rows containing missing values (geom_point).  
  
## Warning: Removed 591 rows containing missing values (geom_point).
```

```
## Warning: Removed 591 rows containing non-finite values (stat_density).
```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 9139 rows containing missing values
```

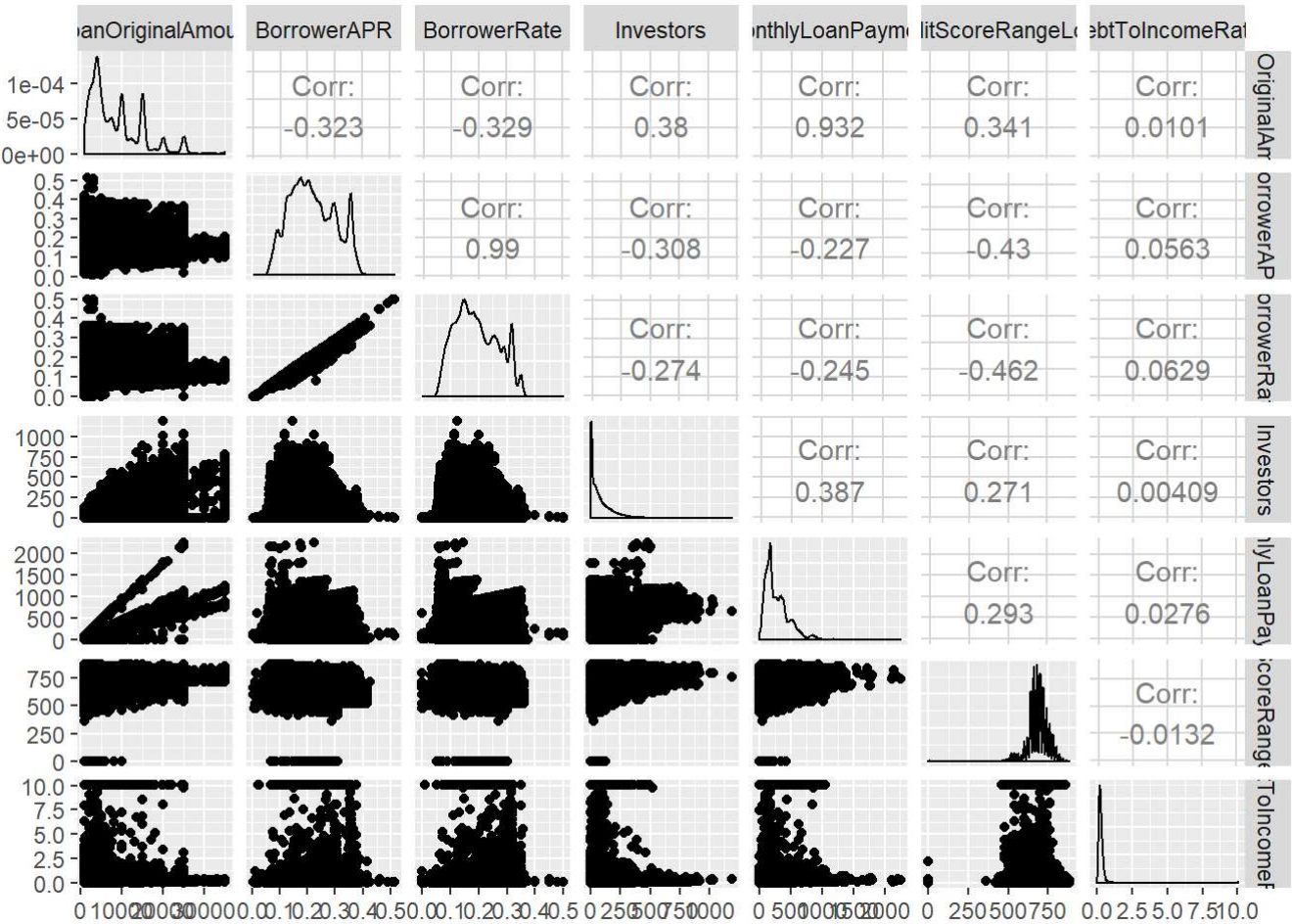
```
## Warning: Removed 8554 rows containing missing values (geom_point).
```

```
## Warning: Removed 8579 rows containing missing values (geom_point).
```

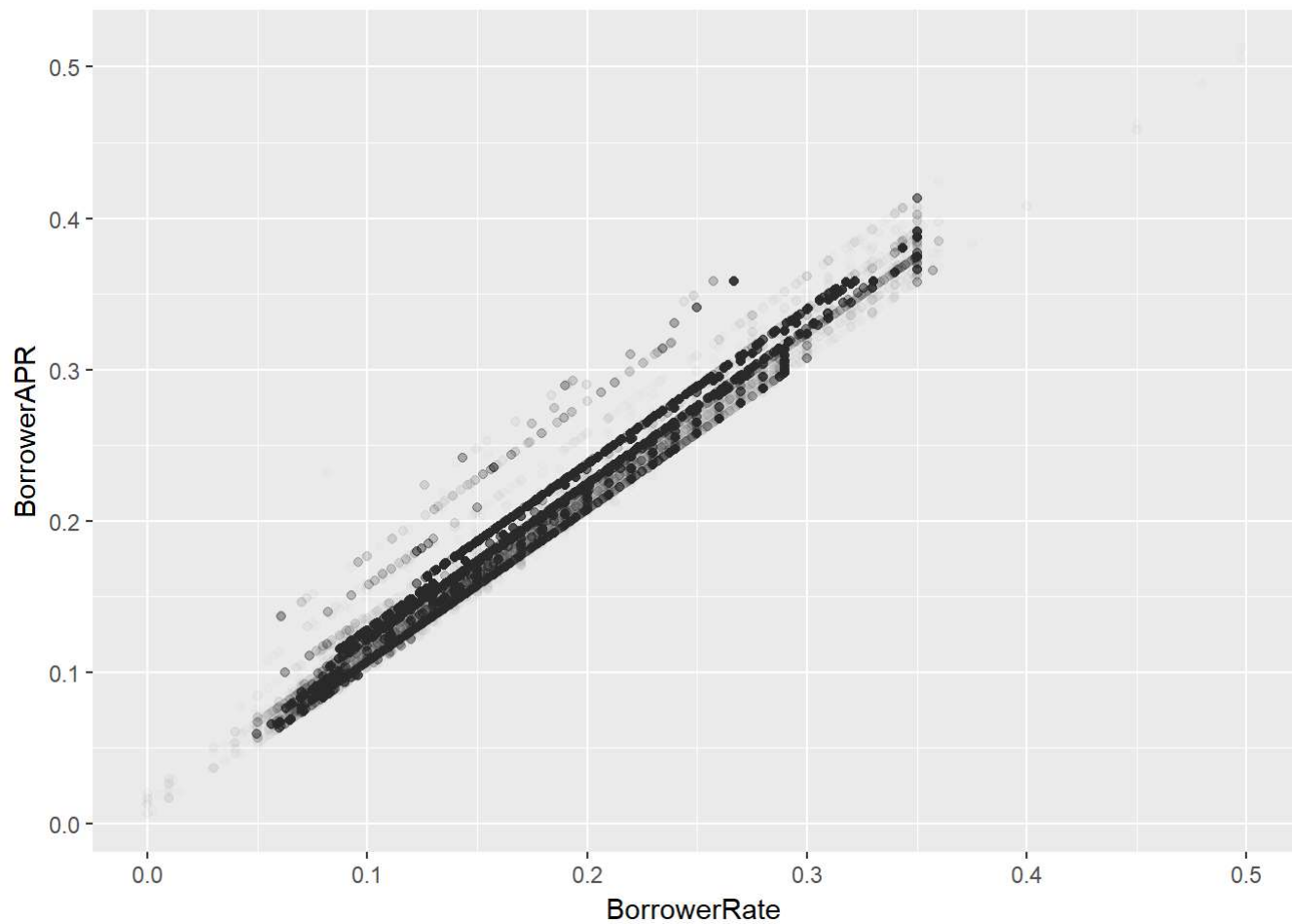
```
## Warning: Removed 8554 rows containing missing values (geom_point).  
  
## Warning: Removed 8554 rows containing missing values (geom_point).  
  
## Warning: Removed 8554 rows containing missing values (geom_point).
```

```
## Warning: Removed 9139 rows containing missing values (geom_point).
```

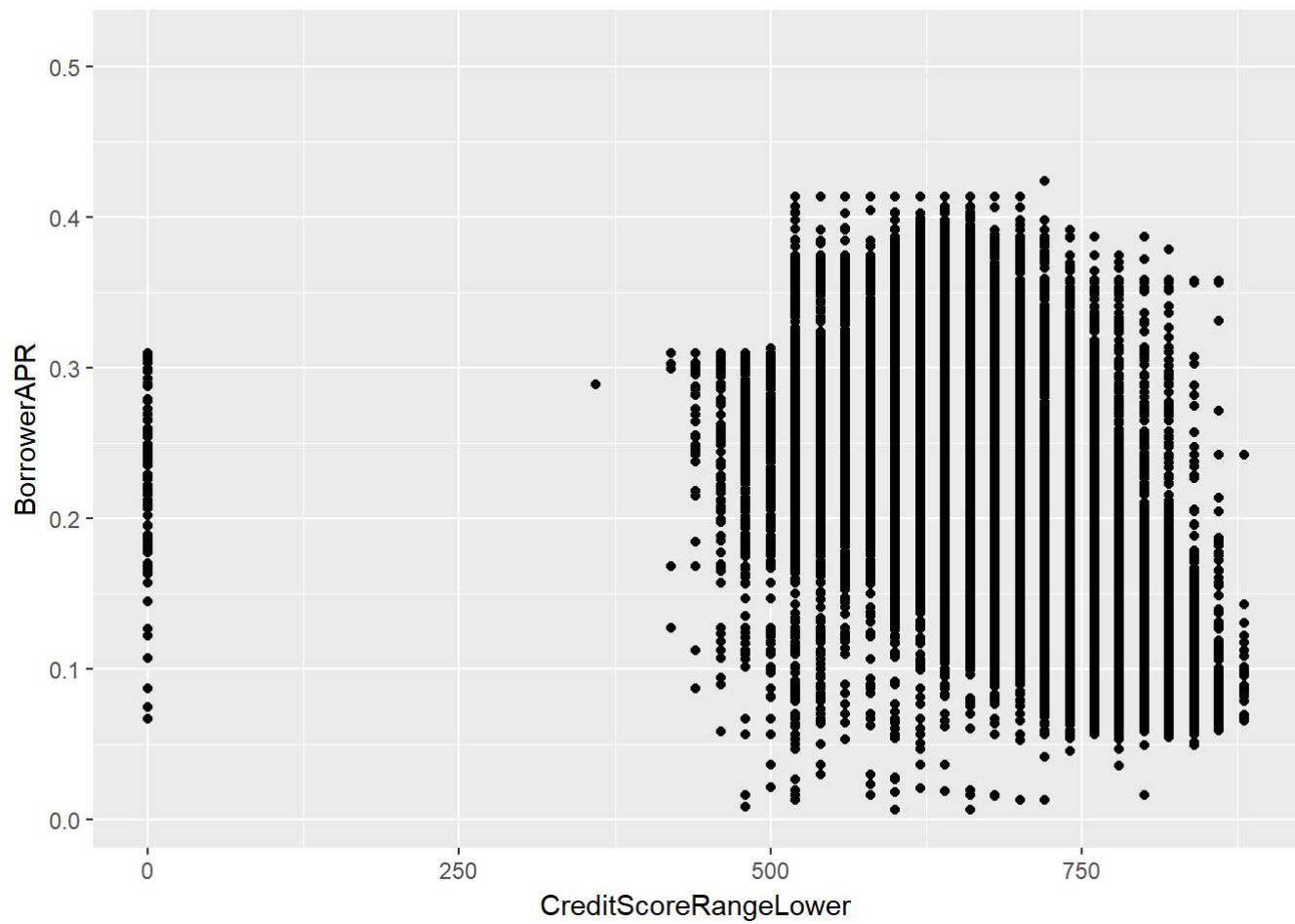
```
## Warning: Removed 8554 rows containing non-finite values (stat_density).
```



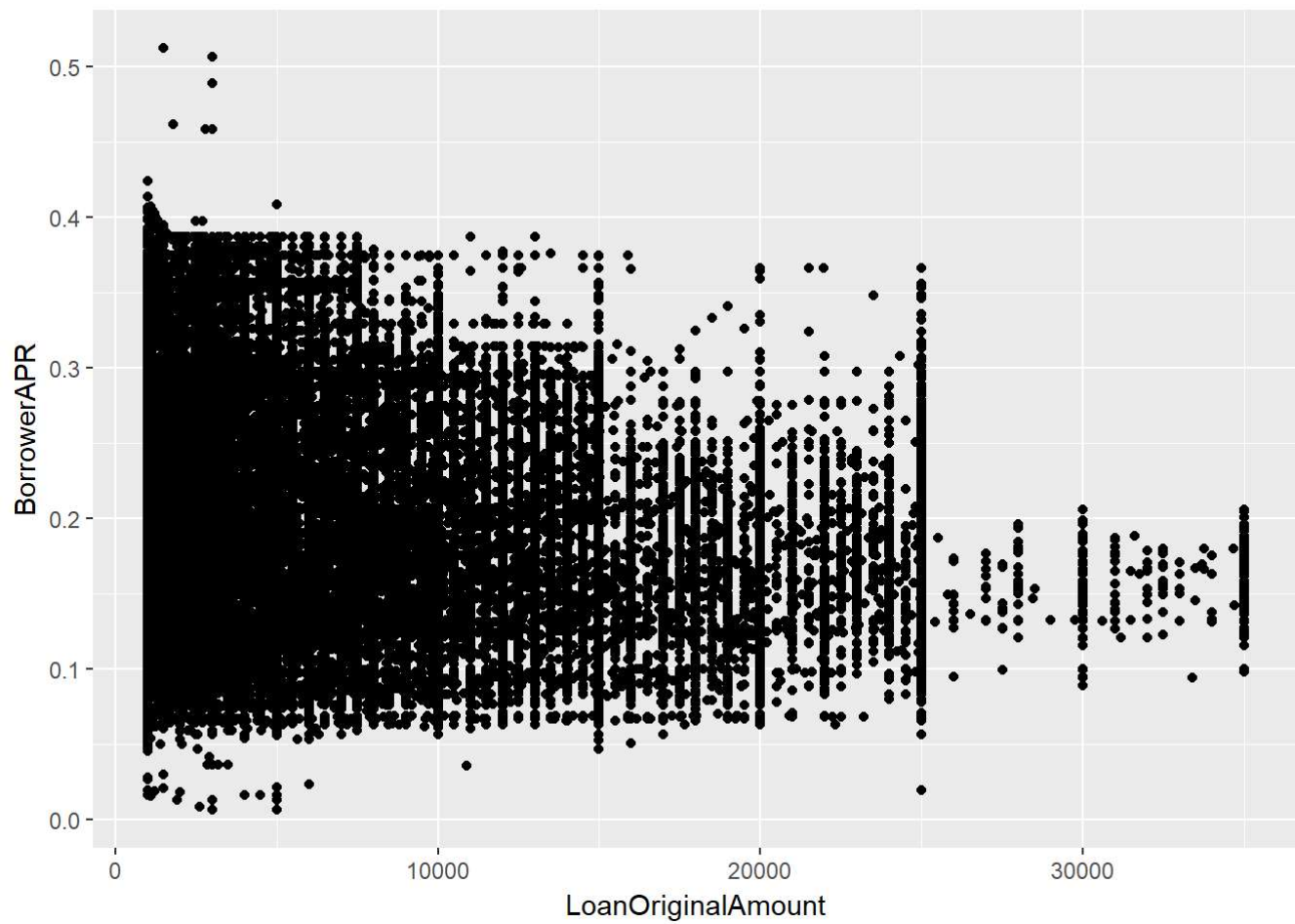
Warning: Removed 25 rows containing missing values (geom_point).



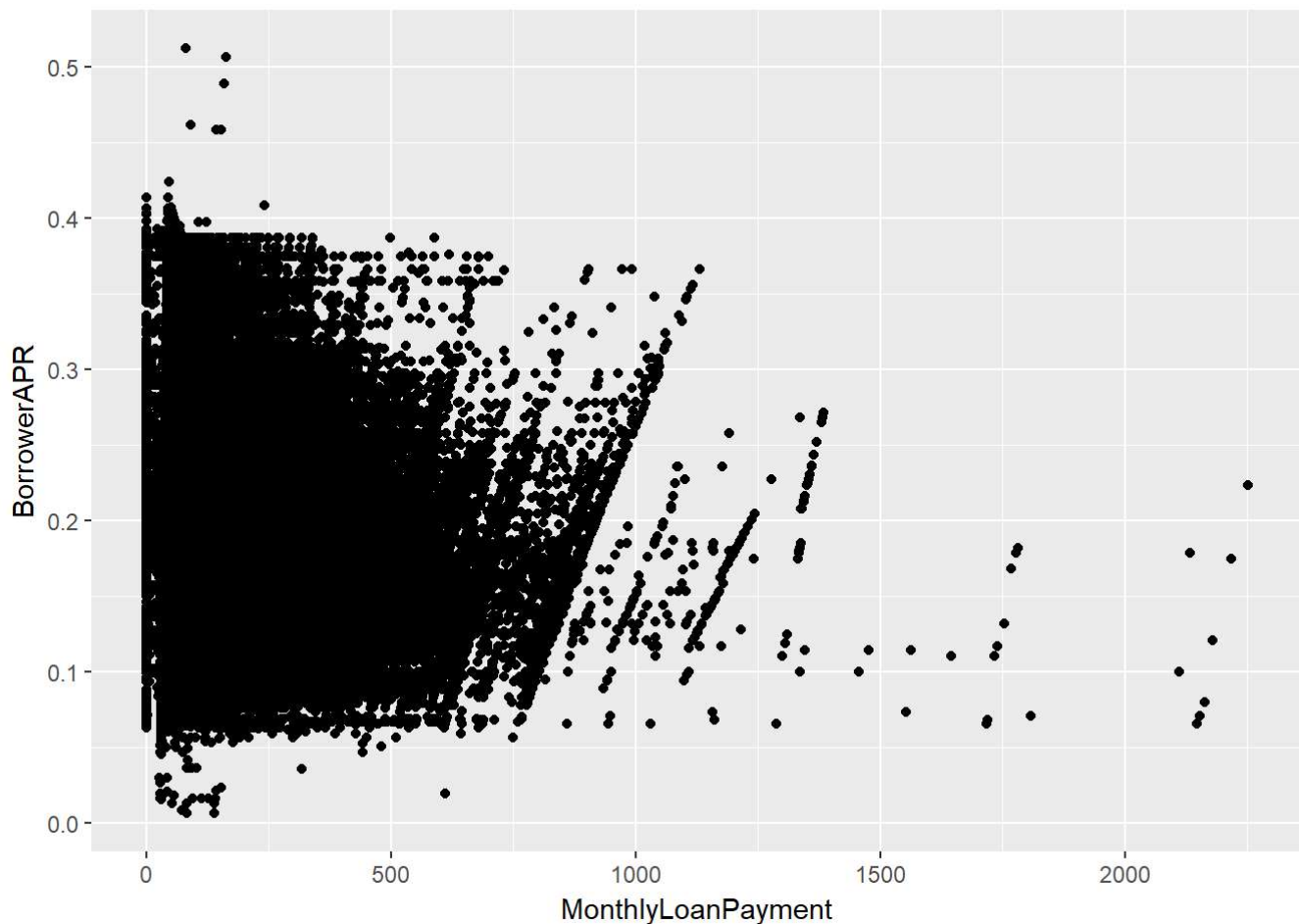
```
## Warning: Removed 591 rows containing missing values (geom_point).
```



```
## Warning: Removed 25 rows containing missing values (geom_point).
```



```
## Warning: Removed 25 rows containing missing values (geom_point).
```



Bivariate Analysis

Tip: As before, summarize what you found in your bivariate explorations here. Use the questions below to guide your discussion.

Talk about some of the relationships you observed in this part of the investigation. How did the feature(s) of interest vary with other features in the dataset?

Did you observe any interesting relationships between the other features (not the main feature(s) of interest)?

What was the strongest relationship you found?

Multivariate Plots Section

Tip: Now it's time to put everything together. Based on what you found in the bivariate plots section, create a few multivariate plots to investigate more complex interactions between variables. Make sure that the plots that you create here are justified by the plots you explored in the previous section. If you plan on creating any mathematical models, this is the section where you will do that.

Multivariate Analysis

Talk about some of the relationships you observed in this part of the investigation. Were there features that strengthened each other in terms of looking at your feature(s) of interest?

Were there any interesting or surprising interactions between features?

OPTIONAL: Did you create any models with your dataset? Discuss the strengths and limitations of your model.

Final Plots and Summary

Tip: You've done a lot of exploration and have built up an understanding of the structure of and relationships between the variables in your dataset. Here, you will select three plots from all of your previous exploration to present here as a summary of some of your most interesting findings. Make sure that you have refined your selected plots for good titling, axis labels (with units), and good aesthetic choices (e.g. color, transparency). After each plot, make sure you justify why you chose each plot by describing what it shows.

Plot One

Description One

Plot Two

Description Two

Plot Three

Description Three

Reflection

Tip: Here's the final step! Reflect on the exploration you performed and the insights you found. What were some of the struggles that you went through? What went well? What was surprising? Make sure you include an insight into future work that could be done with the dataset.

Tip: Don't forget to remove this, and the other **Tip** sections before saving your final work and knitting the final report!