Capstone Project - Slide Deck

Becker

2018-08-03

Introduction

- Wishpond has a high rate of unsubscribed customers
- Predictive models can be used to idenitify those who have a higher chance of unsubscribe and variables that have high correlation with unsubscribing. Thus, offering a recommendation on what to focus on to reduce the amount of unsubscribed customers.

Data

- The database contains a total of 14,417 observations and 21 variables.
- The dataset for this project was acquired from Woopra, a platform that tracks all customer activities on Wishpond platform

Data Limitation

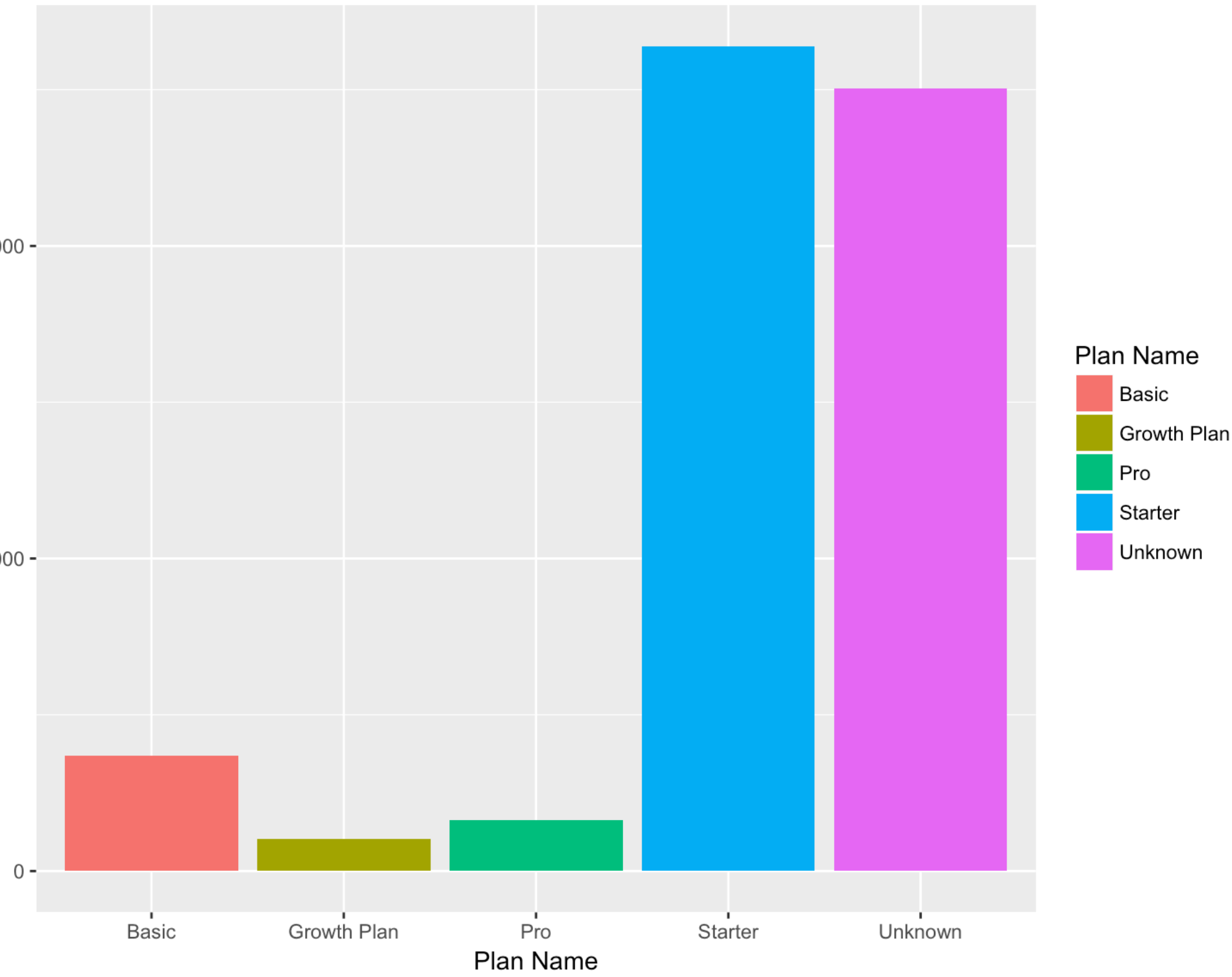
- Large amount of missing data
- Large amount of data from customers that have unsubscribed is lost the moment
- The moment a customer unsubscribe, the amount of leads acquired and amount of tools built goes to zero.

Data Wrangling

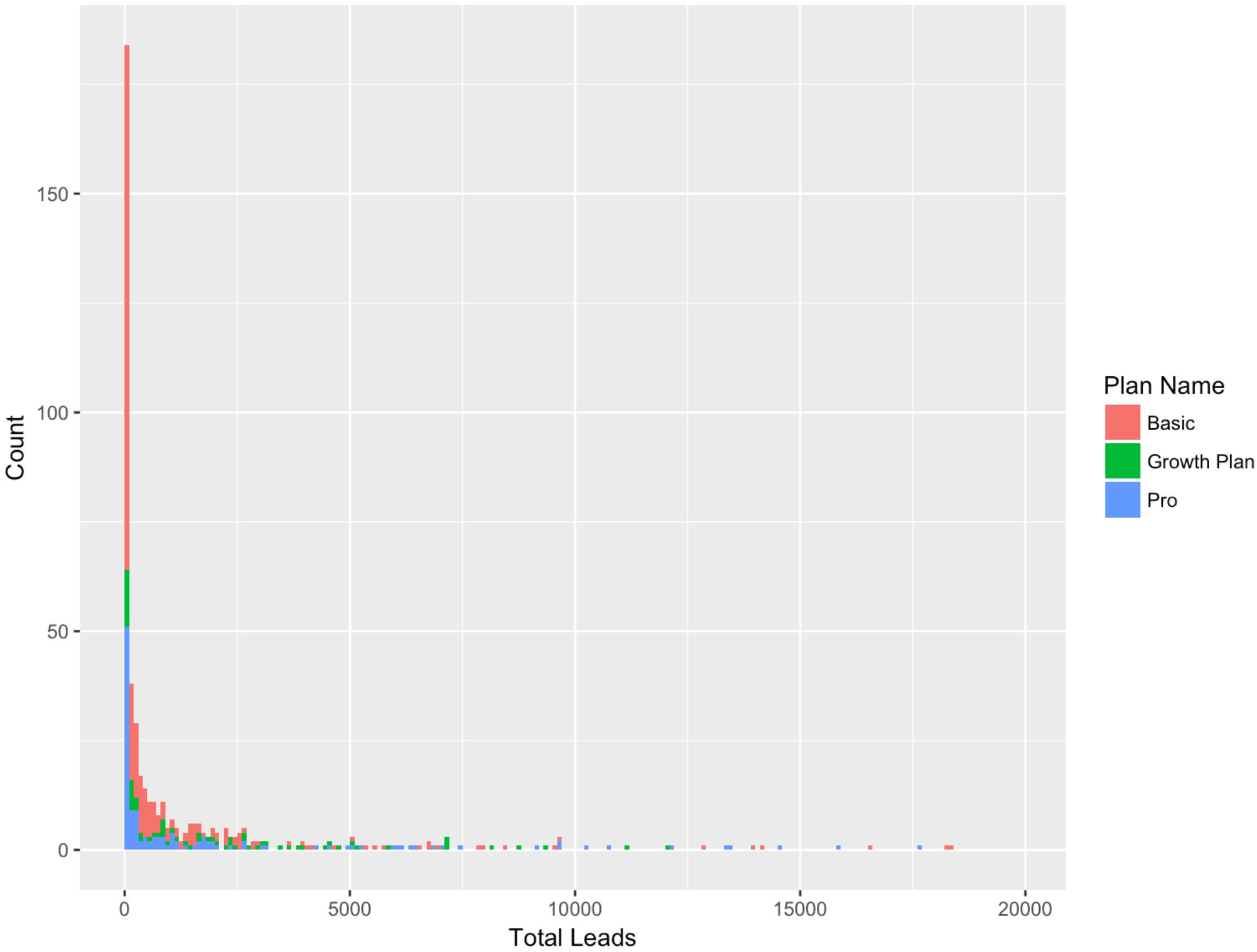
- Lad data from 3 different CSV files and join them into one.
- Data has a lot of missing data
- The NA values were replaced with values that can create a more meaningful and understandable data
- Some variables had a lot of different values that could be identified by a single level.
- Thus, different levels were combined into a single one. (Ex: Different levels of basic plan were replaced with basic plan)
- Variables were renamed and rearranged for better visualization of the database.

Data Visualization

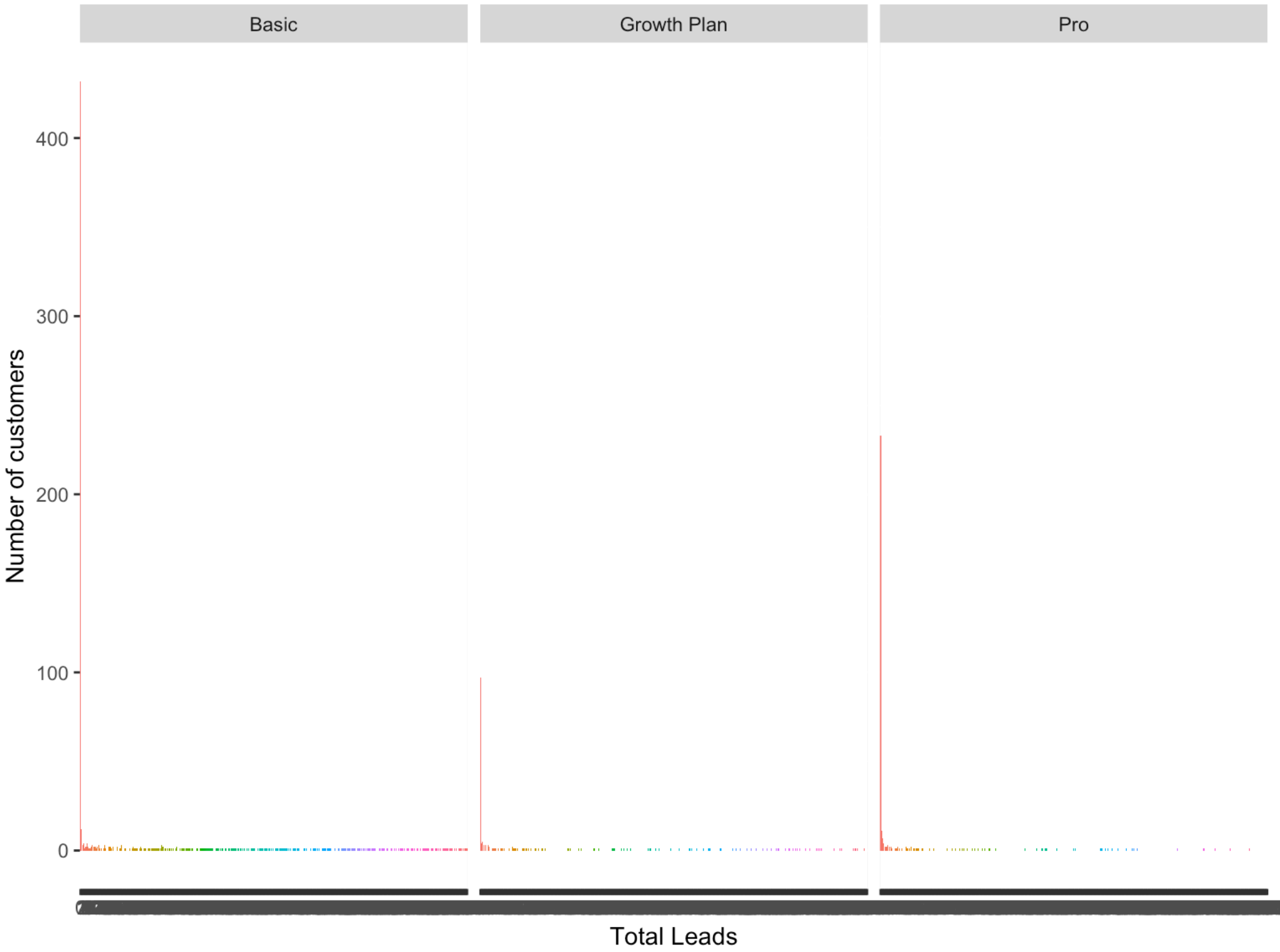
- All Subscribed customers



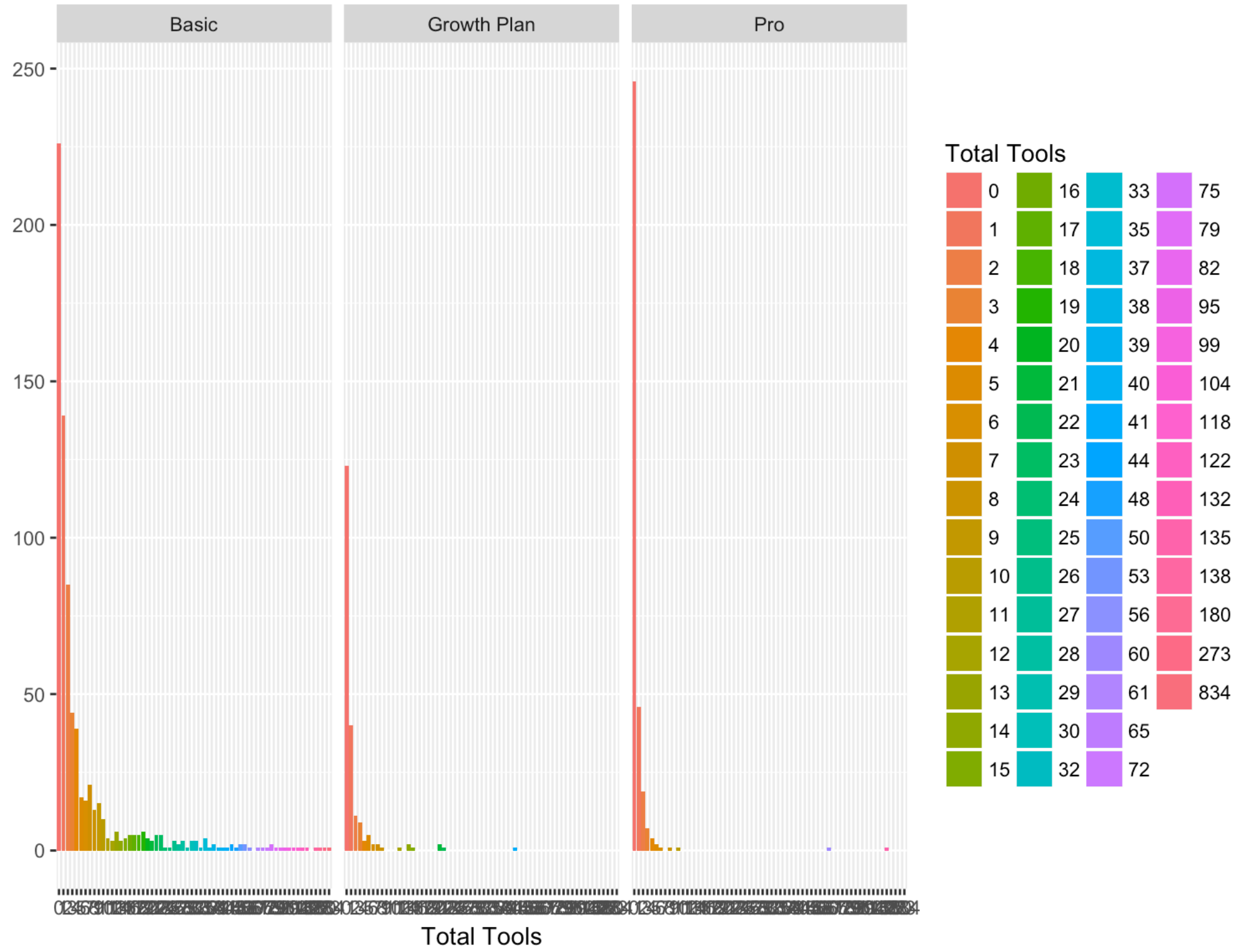
Paying Subscribers



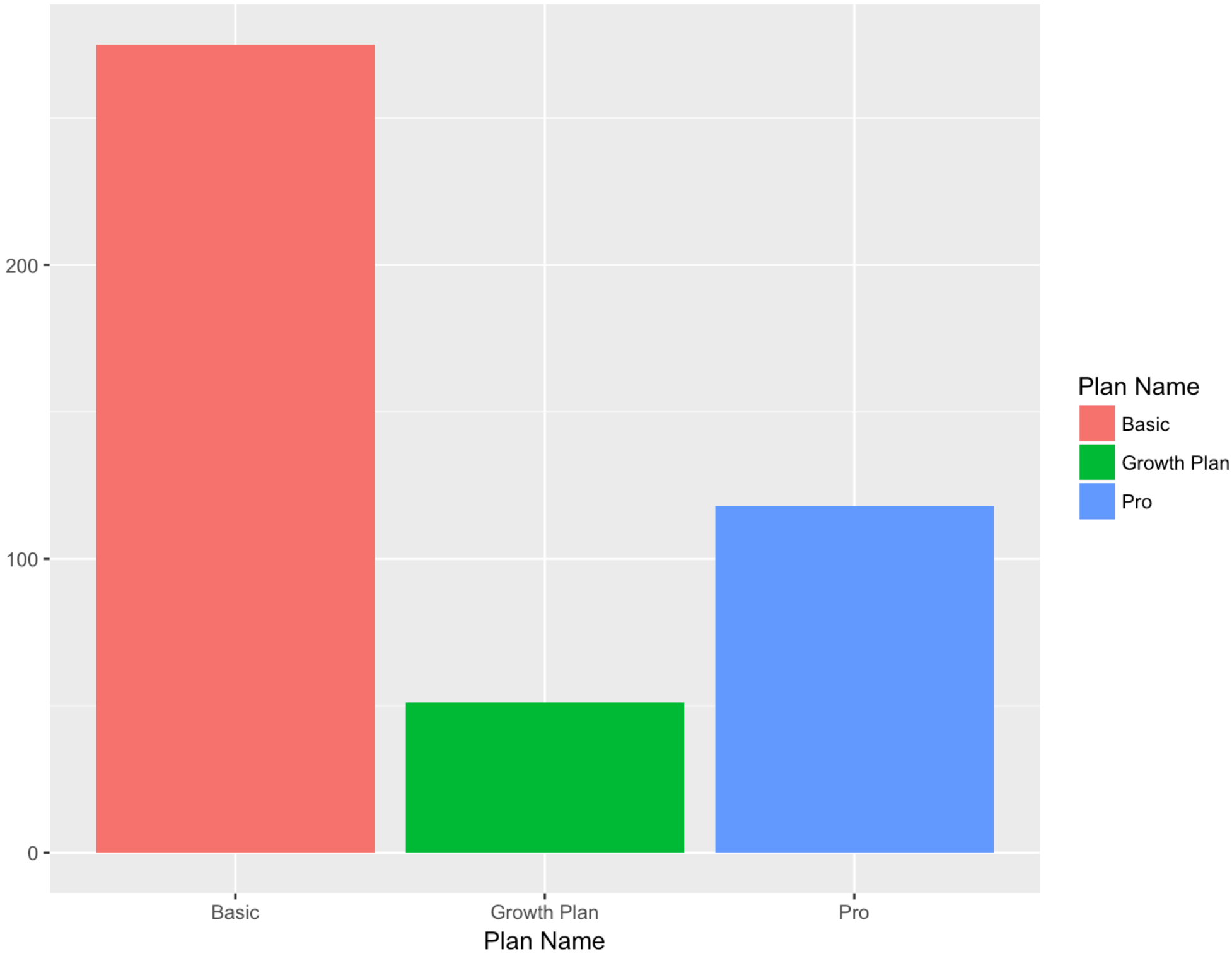
Total amount of leads and their plan



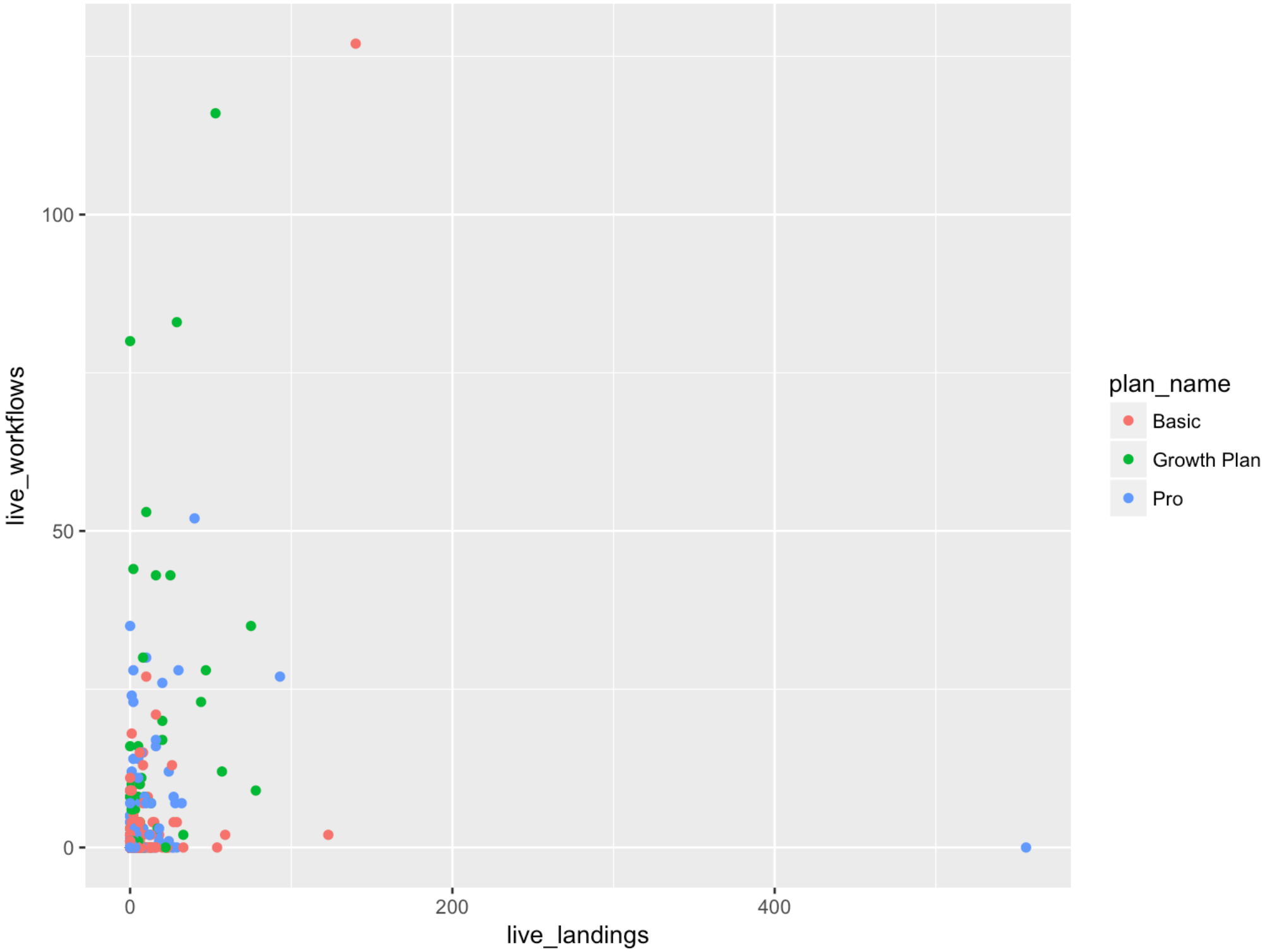
Total amount of tools and their plan



Customers that havent acquired any leads nor created any tools

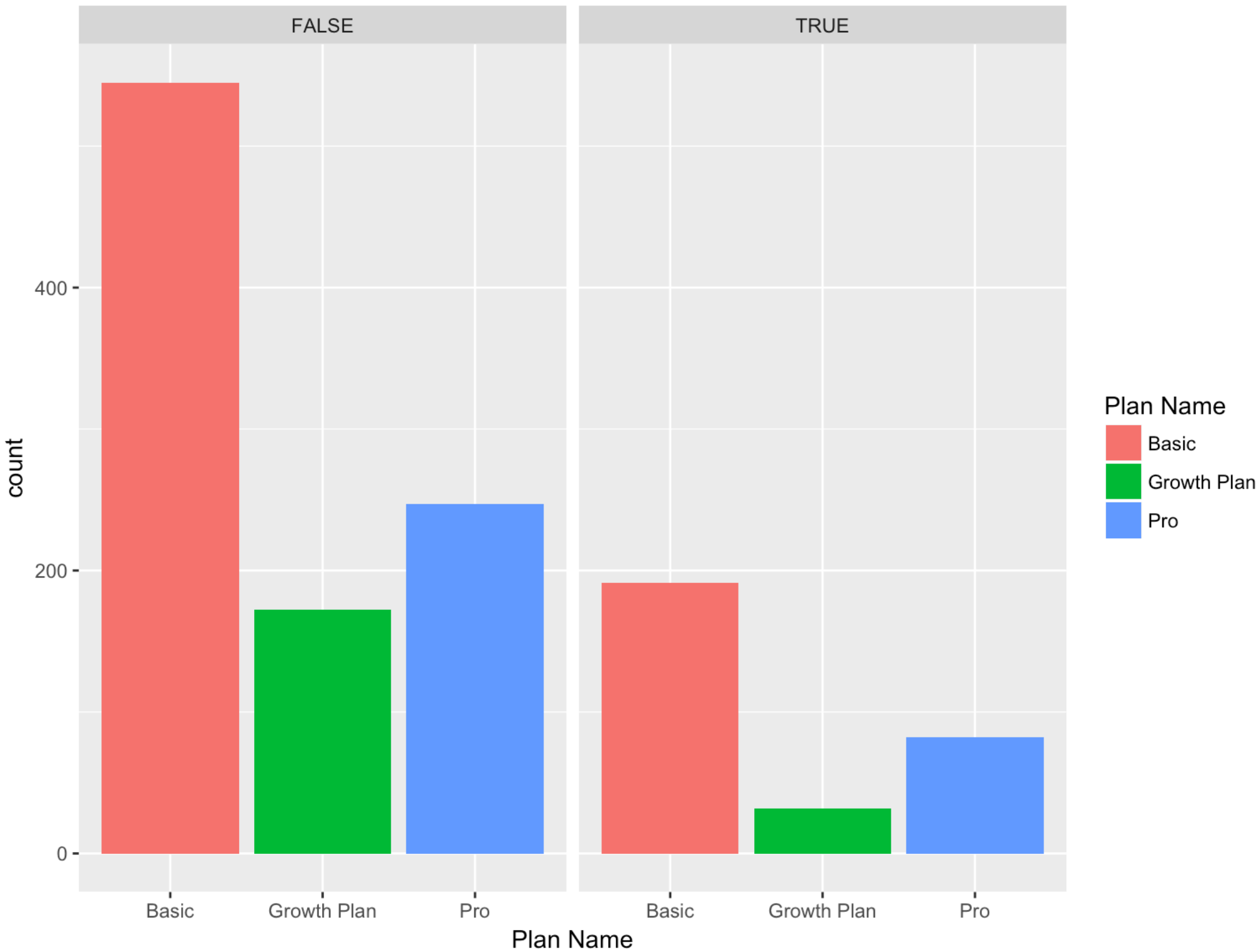


Live Landings Pages & Live Workflows



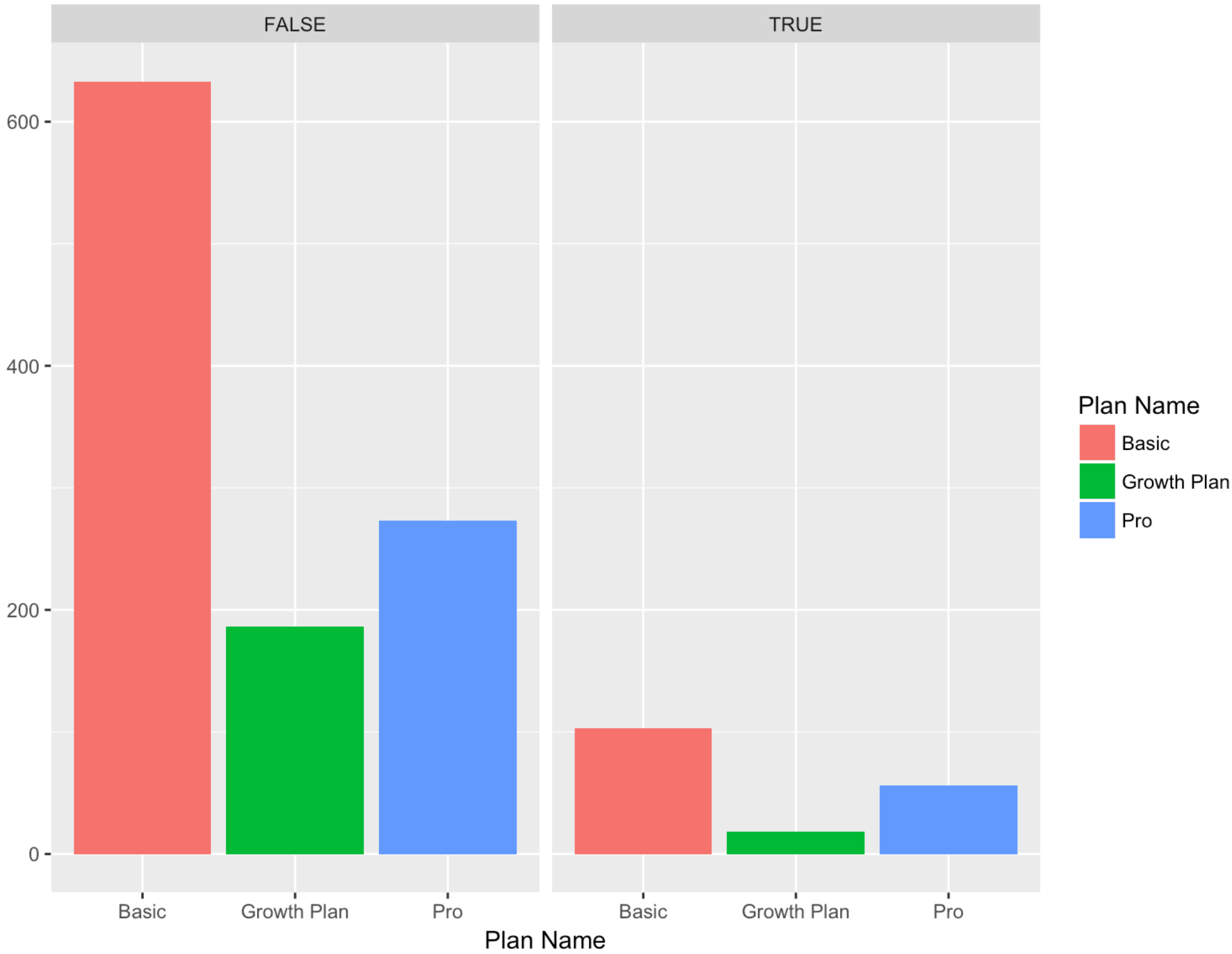
Sales Agents

- Customers' plan & if they came through a demo with a sales agent



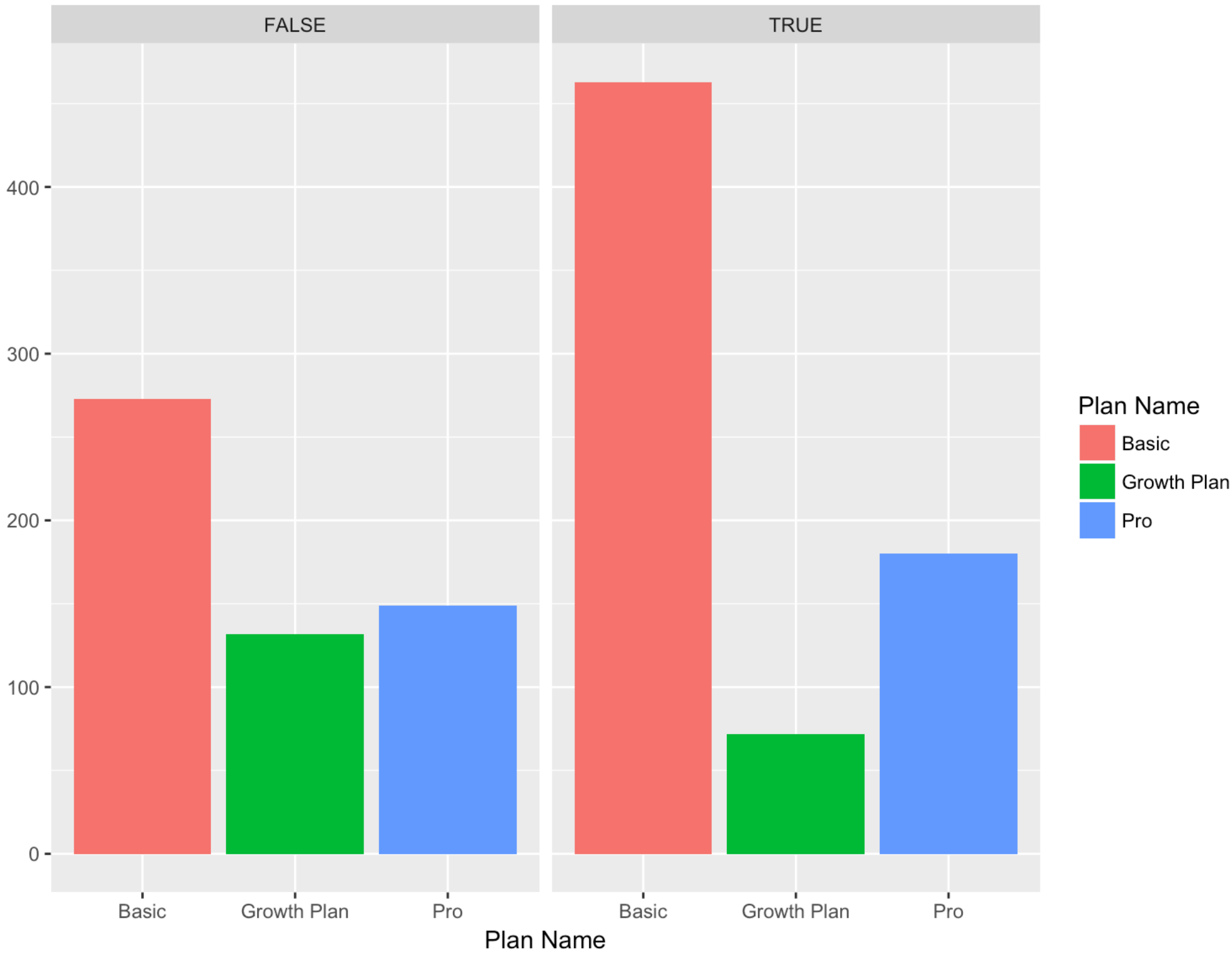
Affiliated

- Customers' plan & if they came through an affiliated link



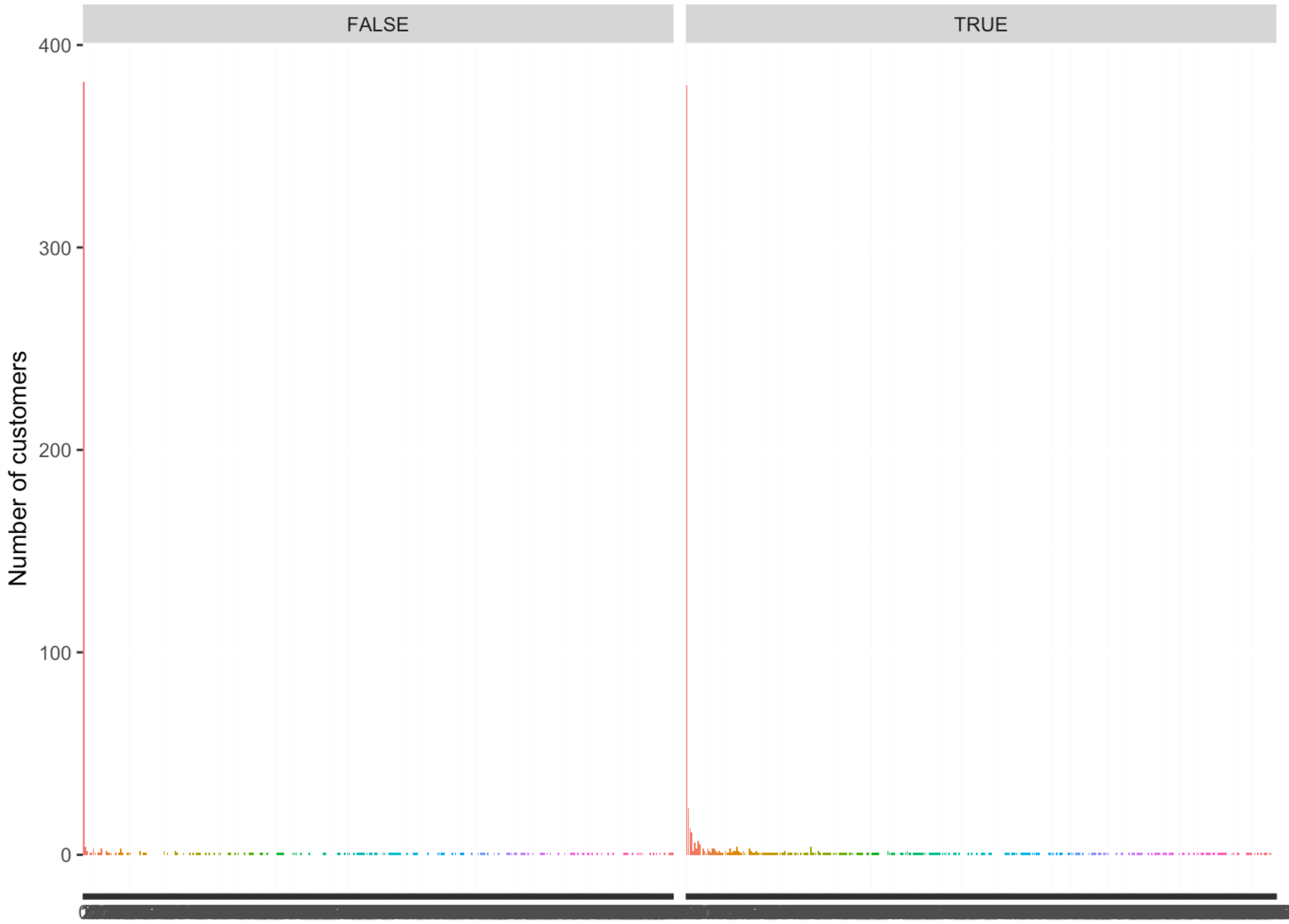
Blog

- Customers plan & if they have read the blog



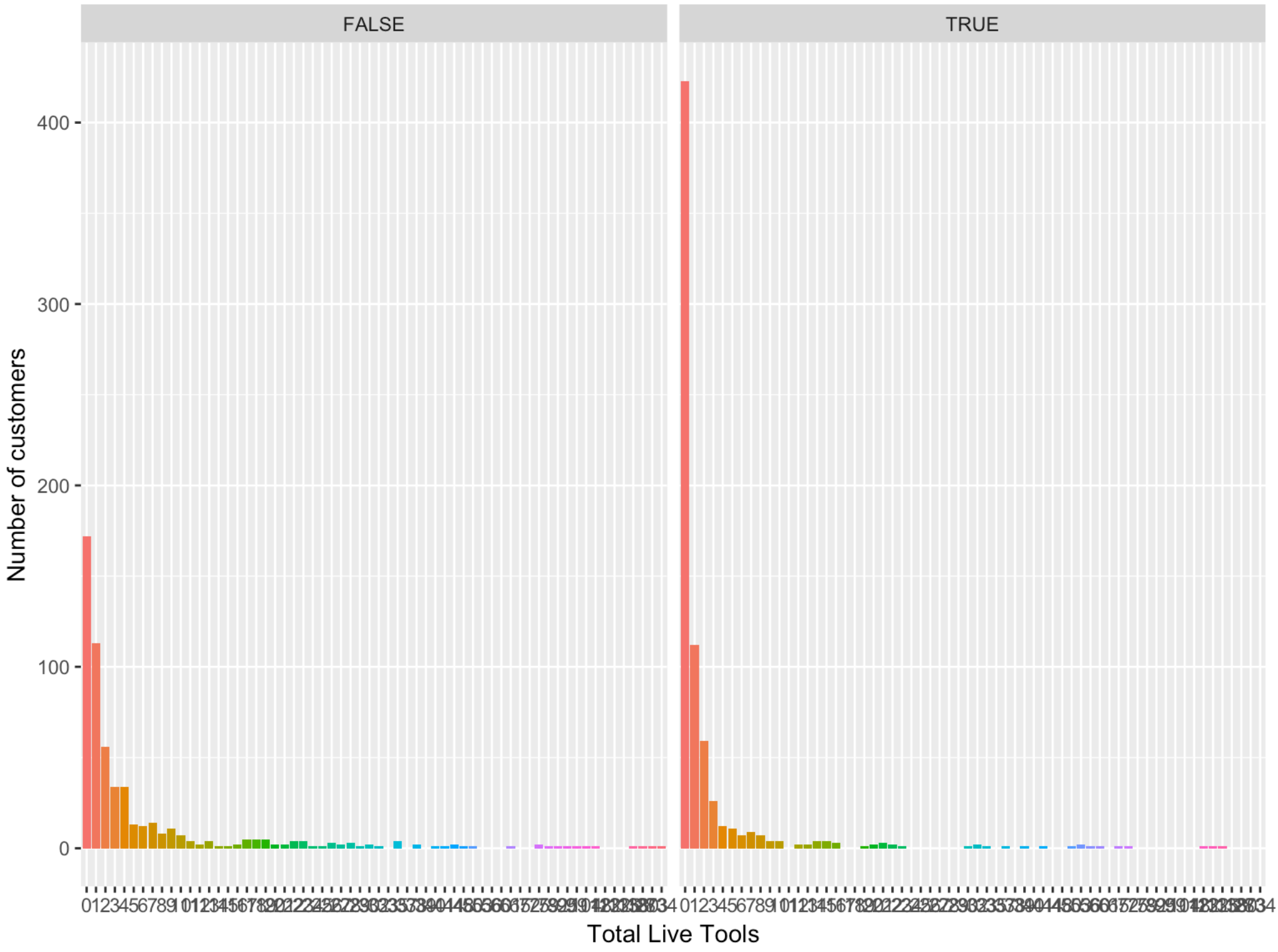
Blog

- Blog impact in lead generation

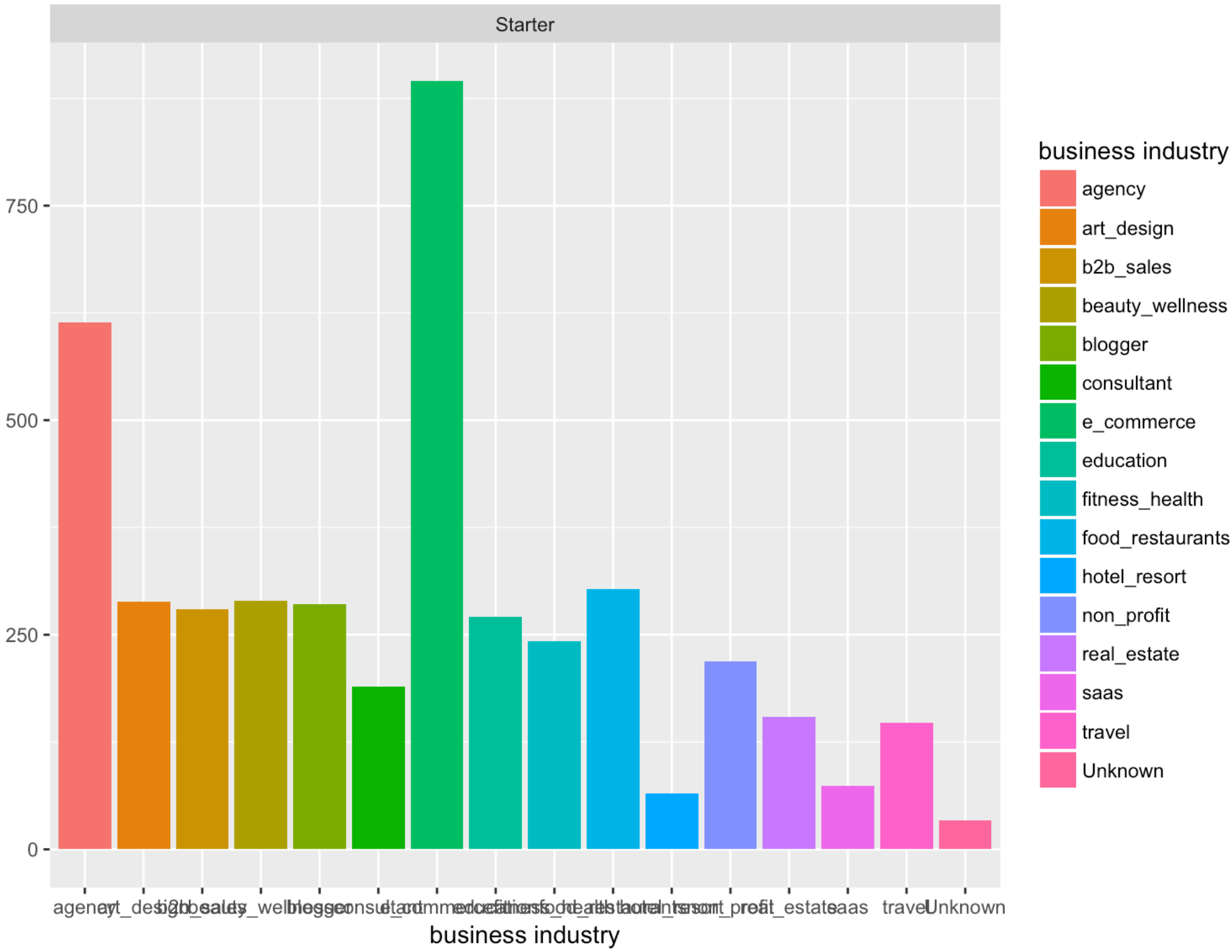


Blog

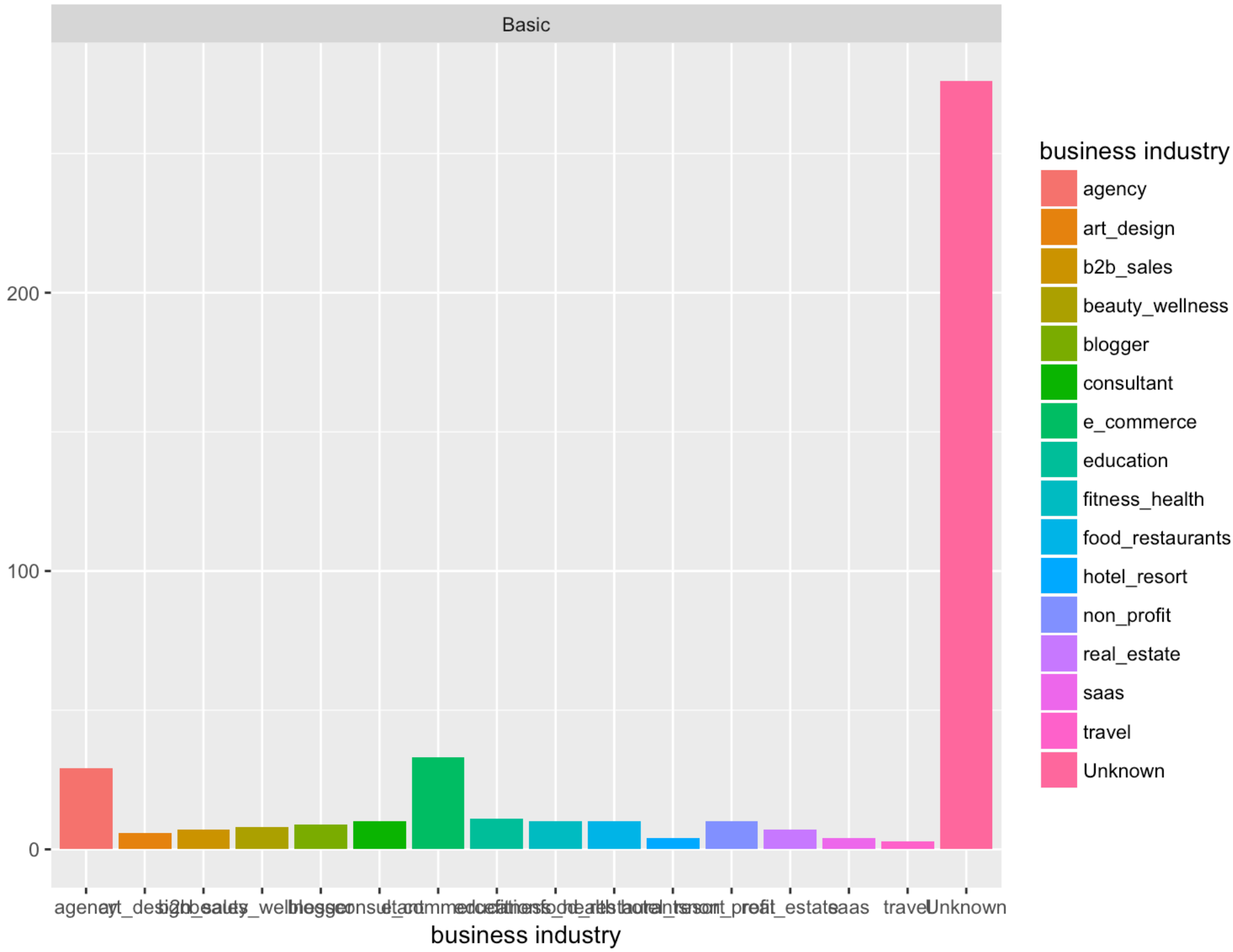
- Blog impact in tools creation



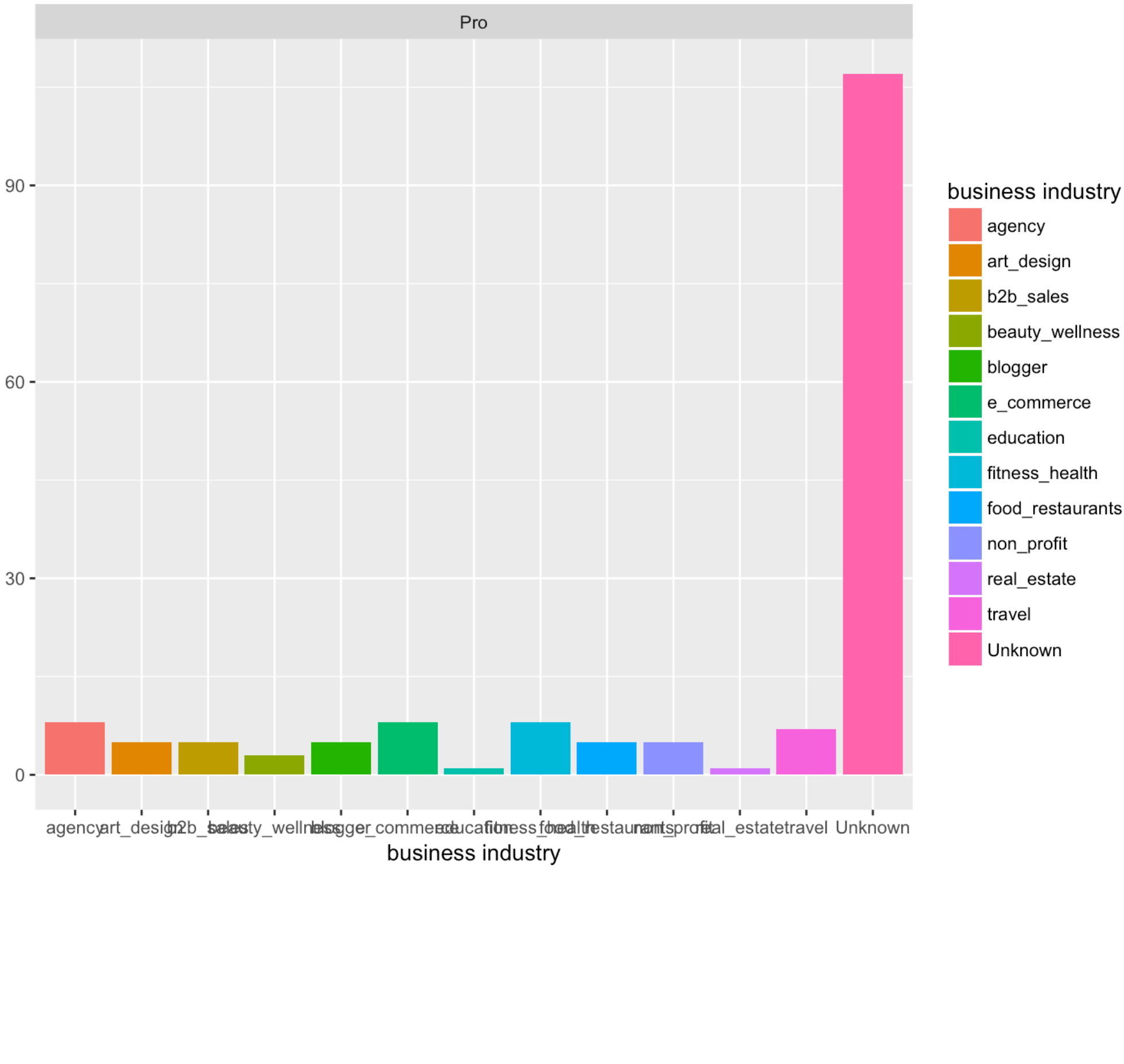
Starter Plan & Business Industry



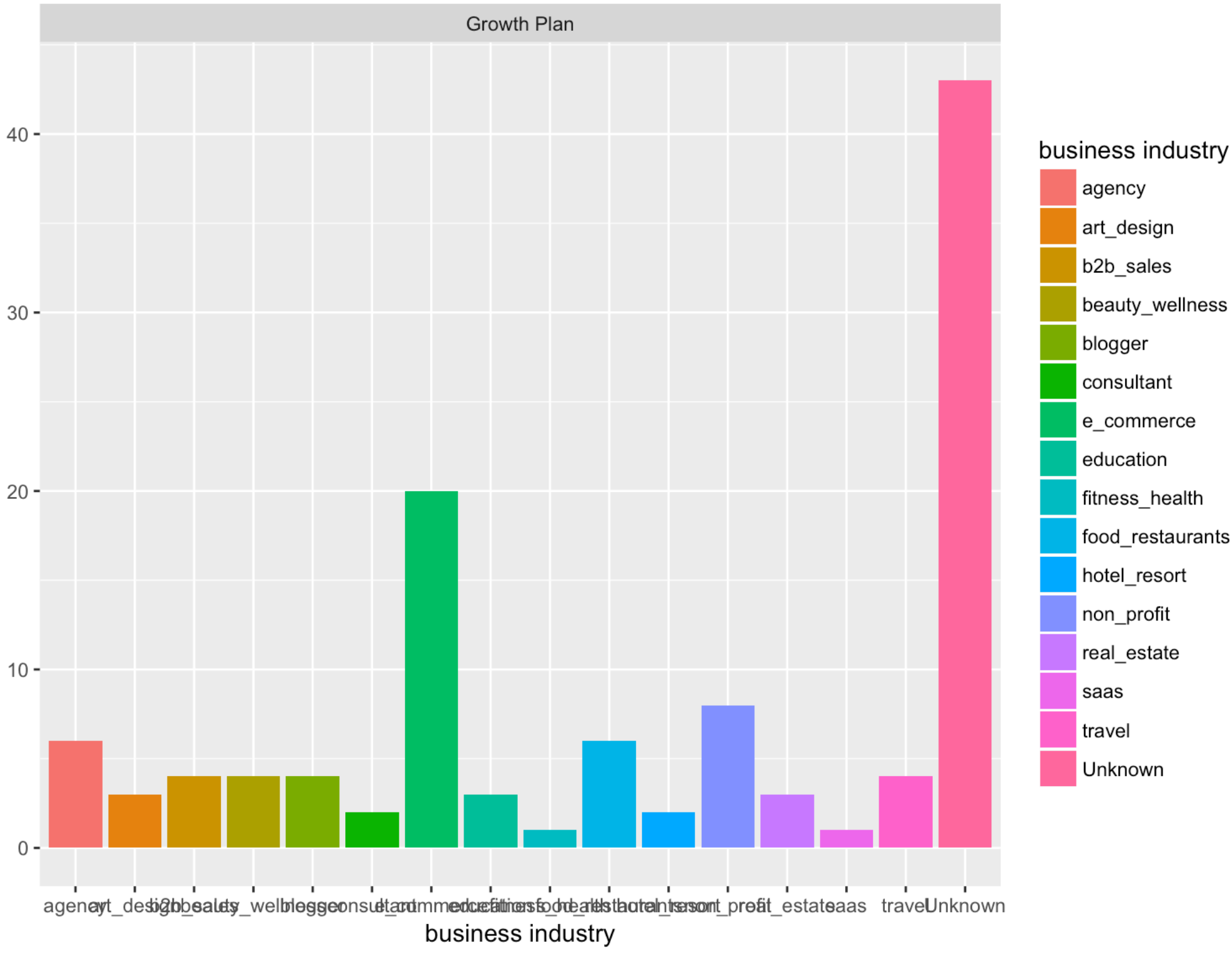
Basic Plan & Business Industry



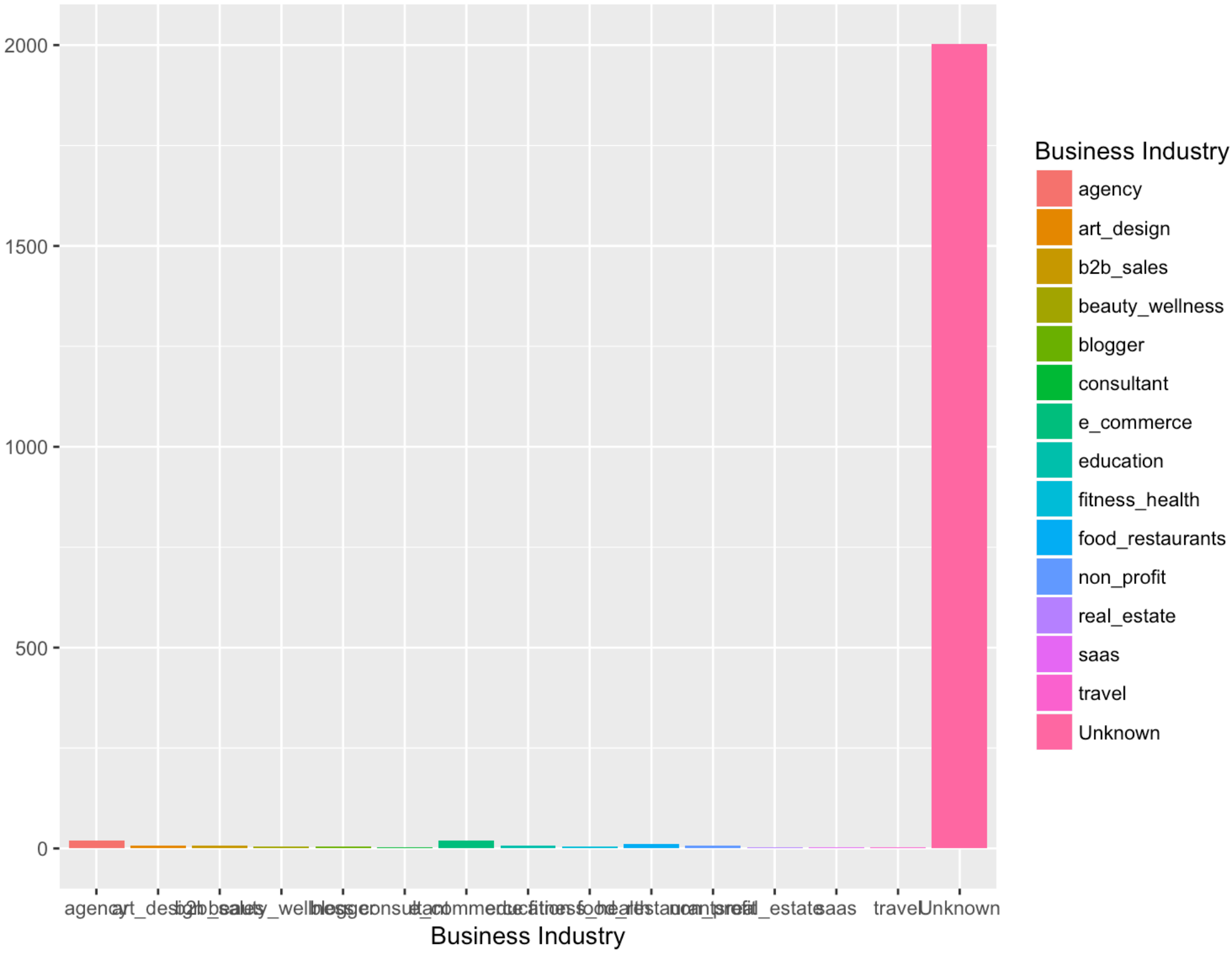
Pro Plan & Business Industry



Growth Plan & Business



Unsubscribed Customers & their business industry



Machine Learning

- Machine learning was used to predict the risk of a customer unsubscribing from Wishpond.
- This is a supervised problem as we are attempting to predict a specific dependent variable using a model based on a set of independent variables.
- This is a classfication problem
- The independent variable is Unsubscribed, FALSE or TRUE.
- This project used different techniques, first it used multiple imputation to replace the missing values with plausible values through probability. Secondly, it used Lasso Regularization to identify the variables that may cause overfitting to our model. Lastly, it used two different predictive models to evaluate which can perform better, logisitic regression and random forest regression.
- To evaluate the predictive models the data will be splitted into two different datasets train (holding larger data) and test (holding a smaller amount of data), in order to provide an unbiased evaluation of the final model.

Machine Learning Findings

- Promissing Variables:
- Website - There is a high correlation that customers that have their own website, tend to stay subscribed with Wishpond.
- Sales - If a customer subscribe after coming from a demo with one of the sales agent there is a significant correlation of them staying subscribed
- Landing page & Pop-Up - There is a high correlation of remaining subscribed if a customer main interest to subscribe to Wishpond is to build a Landing Page or Pop-Up
- Cancel contact support - There is a high correlation that unsubscribed customers have talked to a support agent before unsubscribing.

Machine Learning Findings

- Lasso Regularization: Variables that cause the model to overfit
  - *Business industry*
  - *Business size*
  - *Total leads*
  - *Read blog*
  - *Plan name*
  - *Cancel feeling*

Machine Learning Findings

- Logistic Regression Model
- Accuracy of 74.9%
- Baseline Method of 70.8%
- The accuracy of predicting who have unsubscribed is slightly better thant the basiline method, which means that our model has a significant prediction

Machine Learning Findings

- Random Forest
- Accuracy is 76.1%
- Baseline is 70.8%
- The accuracy of predicting who have unsubscribed is slightly better thant the basiline method, which means that our model has a significant prediction
- Random forest regression produced a better model than the logistic regression by over 1%, with an accuracy of 76.1% while logistic regression model was 74.9%

Machine Learning Findings

- Recommendation

  - *From the machine learning analysis we were able to see that there is a high rate of customers unsubscribing from the platform, and if nothing changes 70% of the customers will continue to unsubscribe from the platform.*

  - *The model identified independent variables that has significant correlation with Unsubscribing. Thus, based on the findings we will provide four recommendations to Wishpoond.*

Machine Learning Findings

- Recommendation
  - *First, increase focus on those who are interested in building Landing Pages and Pop-ups, as both shows high correlation with remaining subscribed to Wishpond. The increase in focus can be done by focusing marketing efforts such as ads and blogs towards people that are interest in landing pages and pop-ups.*

Machine Learning Findings

- Recommendation
  - *Second, marketing efforts should also focus on businesses that have own have own a domain or just purchased a domain. To reach this audience the marketing team could create content on how Wishpond can integrate with their website and paid ads to those that are engaged with domain related platforms.*

Machine Learning Findings

- Recommendation
  - *Third, marketing efforts should not only be focuse on acquiring new leads, but should also focus getting the lead to book a demo with a sales agent. This can be done via email marketing campaign to get them to book a demo and automated SMS to remind leads about their demo.*

Machine Learning Findings

- Recommendation
  - *Last but not least, there was a lot of missing data, and the accuracy results could be improved if data from unsubscribed customers were automatically set to 0. Therefore, we recommend saving the data right at the moment the customer decide to unsubscribe from the platform, giving us more insights on what they did and the reason the might have cancelled.*