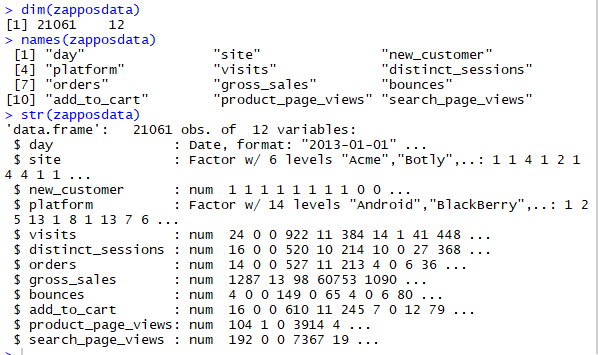
**Univariate Analysis**

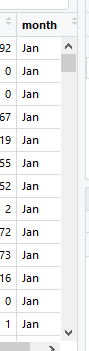
* **Data Observation and Data Cleaning**

Before we start Data Analysis, let’s try to understand the structure of the data.

* First we load the file in the R using the **xlsx** package. After loading the file, let’s try to understand the data by using **names** to explore the names of the columns in the table, **str** to know more about the structure of the table and **dim** to know about the dimensions of the table.



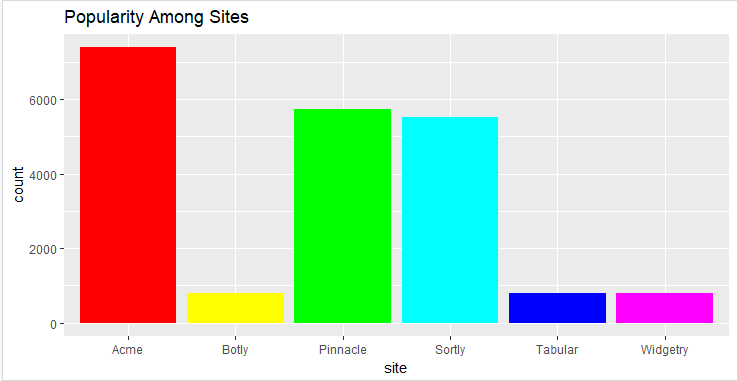
* Adding another column which contains the month, so that we can analyze the data based on the Month.



On observing the above, we find that data for months **March, April and May** is missing.

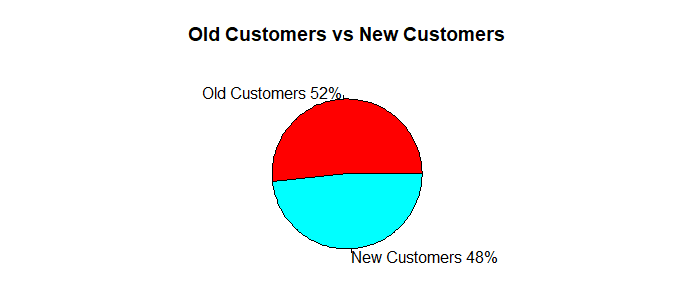
Since we have observed our dataset and cleaned it. We can now move to visualize our data and draw insights.

* **Visualization**
* Let’s try to find out which **site** is **most liked or visited by the customers**. This can be done by carrying out univariate analysis of the Websites.

****

**Popularity of Sites Among Customers**

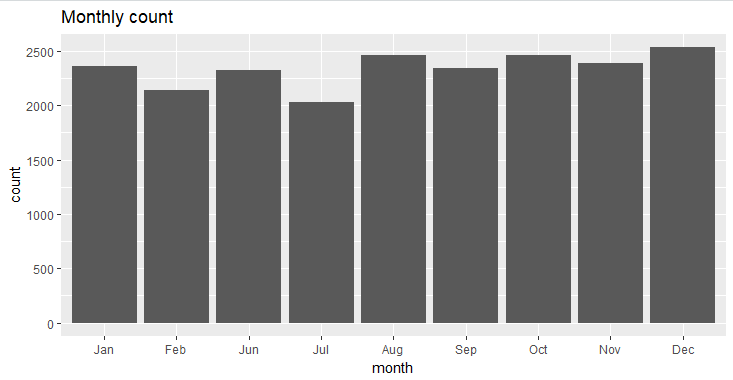
* From the above graph we can say that **Acme** is the most popular site.
* It is highly visited by the customers. We should try to look what’s different in Acme that’s not present in other sites since it is the most popular and try to implement the same features in other sites as well.
* Sortly is the second most popular site. Pinnacle is also not very far behind and is doing fine.
* Botly, Tabular and Widgetry are lagging behind. We should look closely what’s wrong in them and try to increase the popularity of these by making them more user friendly or offering better discounts. We should try and compare the difference between these sites and Acme and try to implement the features of Acme in these sites as well.
* Now, let’s analyze the number of **old customer and the number of new** **customer** so that we know whom should we concentrate on.

****

**Old Customers VS New Customers**

From the above pie chart we can say that the number of old customers is slightly greater than the number of new customers. So, we should give attractive deals to old customers to retain them since they visit our sites regularly. We should try and increase the New Customer base by offering flashy discounts and cool products on our site.

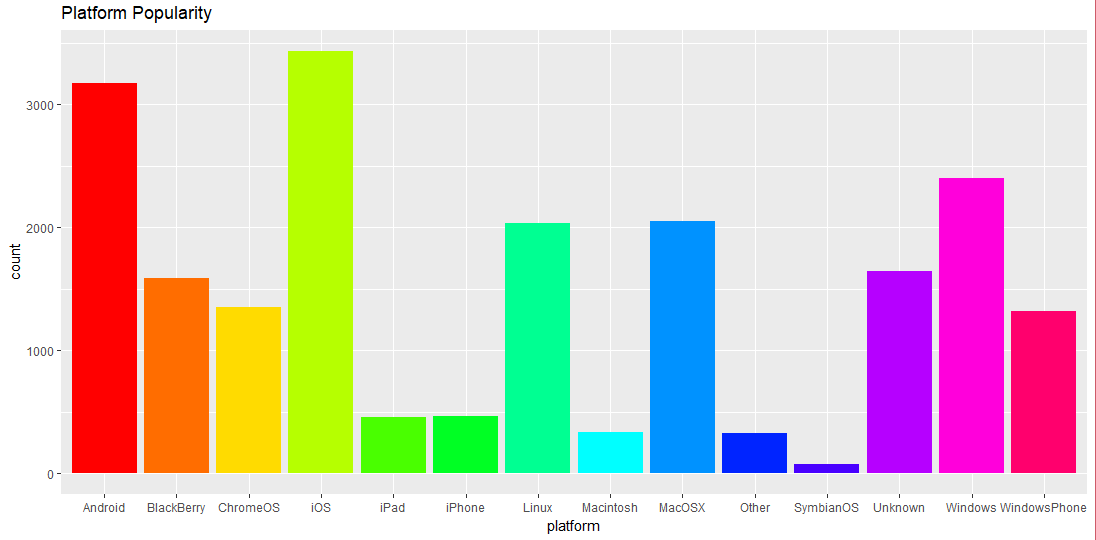
* Let’s see how the customers are divided by month i.e. which month had the most number of customers. This can be seen from the graph below:-



**Monthly Customers**

From the above graph the following inferences can be drawn:-

* The max number of customers were in December. Thus, we should make sure that our site is able to handle more number of customers at that time. We can provide good discounts in December so that our conversion rate is high and we can make profits.
* The minimum number of customers were in the month of Jun so we should try to find out what went wrong or what was different in June and make sure that it doesn’t happen in other months.
* We also observe there is no data for the month of March, April, May.
* Let’s analyze the popularity of the platform i.e. how which platform users most frequently visit our sites. We can understand this by plotting a bar graph which has platform on the x axis and shows the count on the y axis :-

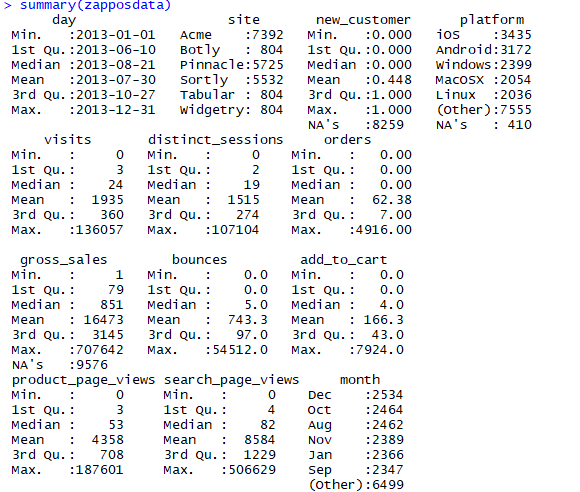


**Platform Popularity**

* From the above graph, we can see that iOS is the most popular platform that people use for browsing our sites.
* Android is the second most popular platform and Windows is the third most popular platform.
* The others like Blackberry, ChromeOS, and Windows Phone have somewhat similar popularity.
* From the above inferences, we should focus on making our websites more iOS and Android compatible as they are the most widely used platforms for accessing our sites.

**Numerical Summary**

We can find the numerical summary of the dataset by using the Summary function.



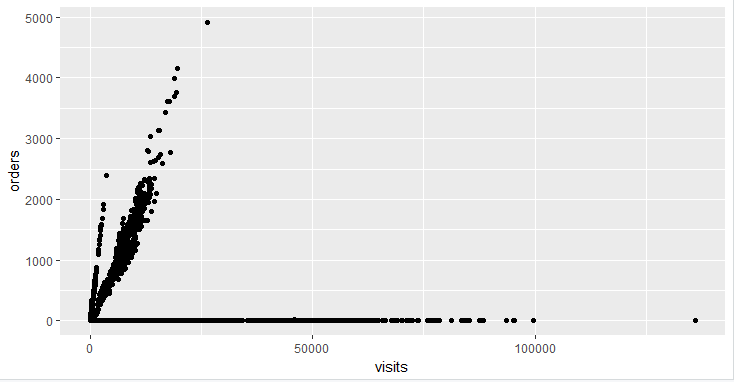
**We see that the numerical summary matches with what we inferred from our visualizations.**

* **The most popular website is Acme according to the visual and numerical data.**
* **The number of old customers is more than the number of new customers**
* **December is the most popular month.**
* **iOS is the most popular platform for accessing our sites.**

**Bivariate Analysis**

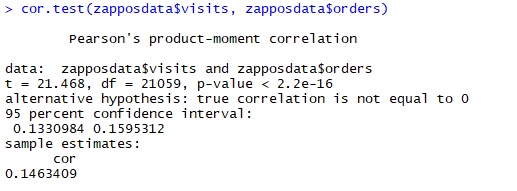
Now, let’s delve deeper and carry out bivariate analysis.

* Let’s try to derive a relation between the number of visits and the number of orders. This will help us know how many people who visit the website actually order something.



**Orders and Visits**

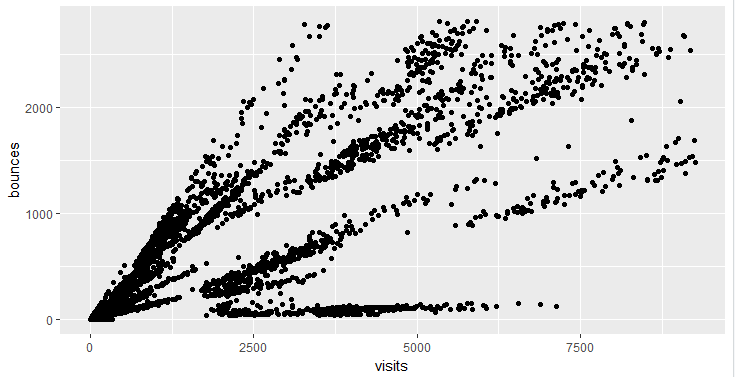
Let’s also find the correlation between the orders and visits using **cor.test,**

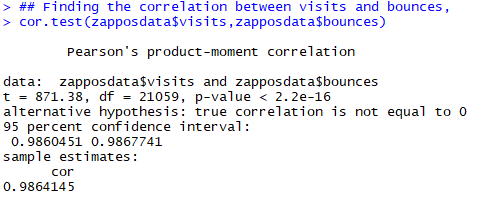


From the above numerical and visual representations, we can say that they are **not strongly correlated** i.e. there are many people who visit the site but do not order. Then according to the graph some people visit the website and order materials as it shows a linear relation for some part of the graph.

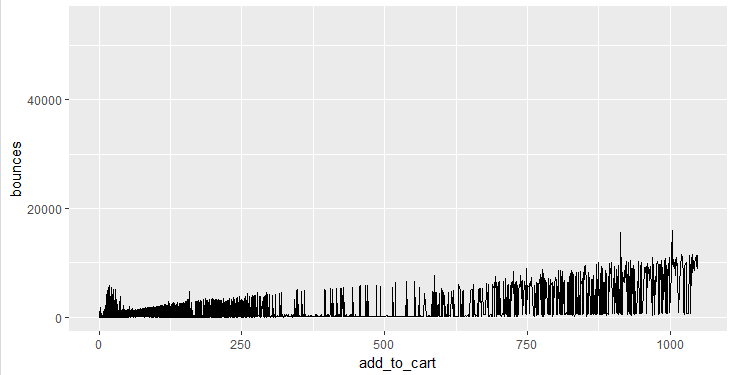
There are some outliers after 100000 visits, we should ignore them as they are far away from the mean.

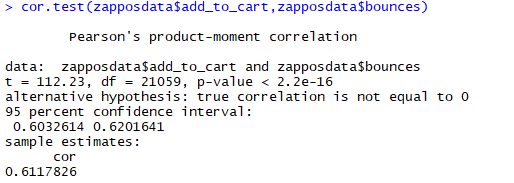
* From the above we know that many people visit the website but some buy and others leave the site. Thus, we should find the relation between the bounces and visits.



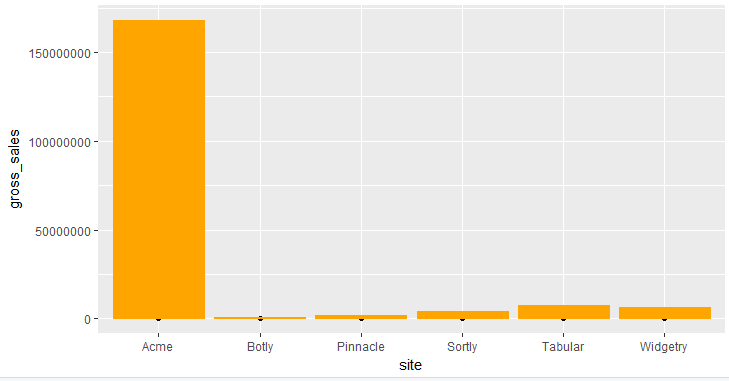


* From the graph we can see that, many people who visit the site leave it. The bounces and visits have a correlation of **0.9864**. This is not a good thing for our sites. We should focus on where we are lacking. We need to delve deeper to know the root cause of the high bounce rate.
* Let’s try to find the relation between the add to cart and bounces.





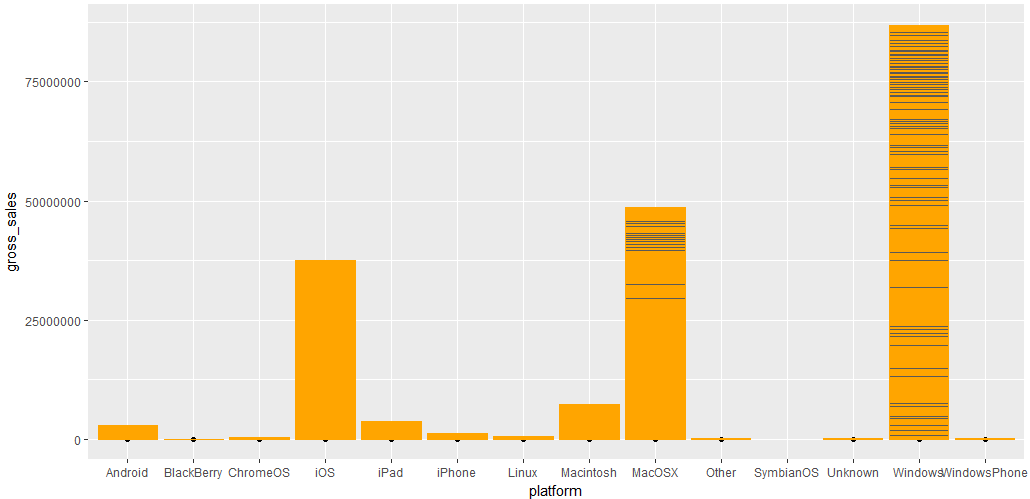
* From the above we can say that bounces and add to cart are related to some extent i.e. it happens that people add products to the cart and then leave the site. This is very bad for our profits and hence, we should make sure that the payment gateway is working well and the UI is easy enough for the user to understand about how to make the payment.
* Let’s try to find which **website** has the **highest revenue**.



**Highest Revenue Website**

* From the above graph we can see that Acme is the clear winner. It generates the maximum gross sales.
* The other websites generate very less revenue in comparison to Acme. Thus, we should think about how to increase revenue from other sites.
* Let’s try to understand which platform generates the maximum revenue.

According to the graph below, Windows generates the maximum revenue. The others lag behind. We should check if our sites are working aptly on the remaining platforms. We should ensure that the UI design and the other functionalities are up and running in the other platform.



**Highest Revenue Platform**

Thus, from the above univariate and bivariate analysis we have been able to draw meaningful insights from our data!