# Loan Summary Statistics

### Introduction

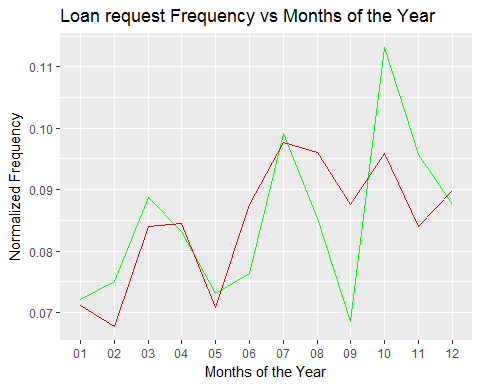
The goal of this project is to create a set of charts that examine who, when, and where the ideal conditions are for lending club to loan money to an individual. With a large Company like lending club there are a multitude of various factors at play from what state does the individual reside in to how many loans they have already given out in a given period. Since we are observing from an outside perspective of the company we have no way of being sure if there are quotas set for employees or other nuances to this companies particular business model, so it will be interesting to examine these factors in depth.

in this report we will be using 3 types of charts, Acception vs rejection, a chart with a pair of lines that shows the data from loans that were accepted against loans that were rejected; percentage difference, a chart that is ment to quantify the differences in terms of percentages between the two groups and to make it easier to see the differences between the two; and heat maps, to show differences over geographical regions in our case the lower 48 states.

### Loan Frequency vs. Months of the Year:

Frequency plots based on twelve months of the year, and number of loans taken out during each month. months with less than 31 days we found the mean frequency per day, scaled it to 31 days and then normalized them based on the total frequency under Loans accepted and Loans Rejected

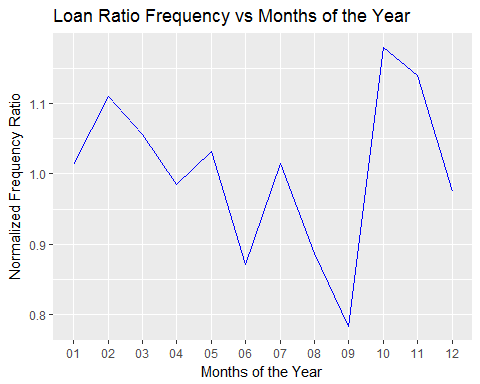
ggplot()+   
geom\_line(data=LAF\_MoTY, aes(x=V1, y=Norm, group=1, colour='Accepted Loan'), color='green')+  
geom\_line(data=LRF\_MoTY, aes(x=V1, y=Norm, group=1, colour='Rejected Loan'), color='red')+  
 xlab('Months of the Year') + ylab('Normalized Frequency') +  
 ggtitle('Loan request Frequency vs Months of the Year')



In the chart above, the green line represents the Normalized number of loans accepted per month while the red line represents the normalized number of loans rejected per month.

For a consumer interested in taking a loan out from lending club we can see that september has the highest frequency of loans accepted out of any month, while august on the other hand has the lowest number of loans accepted per month.

ggplot()+   
geom\_line(data=LDF\_MoTY, aes(x=V1, y=Norm, group=1, color='Accepted Loan'),color='blue')+  
 xlab('Months of the Year') + ylab('Normalized Frequency Ratio') +  
 ggtitle('Loan Ratio Frequency vs Months of the Year')

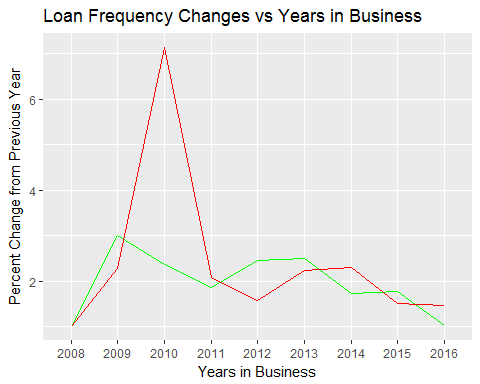


The above chart is created to confirm our hypothesis from the two lines in the first chart on Loan frequency. In september there is roughly a 20% increase in loan acceptances over loan rejections over all. and in total we can see there are 7 / 12 months that have a higher rate of loans being accepted than loans being rejected.

### Percent Change in Loan Frequency vs. Business Year:

Loan Frequency by Year; These group of charts plots the percentage increase in the # of loans from the previous year.

ggplot()+   
geom\_line(data=LAF\_Year, aes(x=V1, y=N, group=1, color='Accepted Loan'), color='green')+  
geom\_line(data=LRF\_Year, aes(x=V1, y=N, group=1, color='Rejected Loan'), color='red')+  
 xlab('Years in Business') + ylab('Percent Change from Previous Year')+  
 ggtitle('Loan Frequency Changes vs Years in Business')

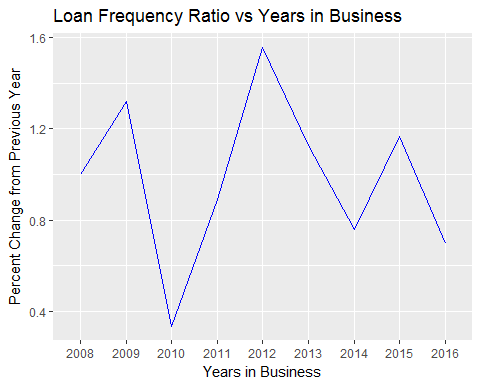


In the above chart the green line represents the percent increase in number of loans accepted over the previous year, and the red line represents the number of loans rejected over the previous year.

In years where the % increase in number of loans accepted is higher then the number of loans rejected we could conclude that the lending club business is doing very well since it must be growing to take out those loans, and therefore would be more likely to accept a loan than other years.

An interesting note with the above is the extremely high number of loans rejected between the years 2009 - 2011, which are considered some of the worst years in the great recession.

ggplot()+   
geom\_line(data=LDF\_Year, aes(x=V1, y=N, group=1, color='Accepted Loan'), color='blue')+  
 xlab('Years in Business') + ylab('Percent Change from Previous Year')+  
 ggtitle('Loan Frequency Ratio vs Years in Business')

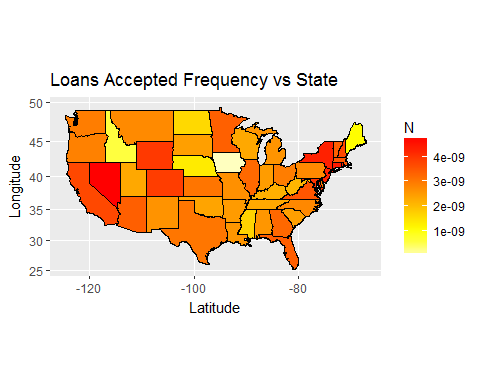


The above chart represents the percent growth in accepted loans divided the percent growth in rejected loans. We can see that at its worst lending club increased its number of accepted loans at a rate of 40% compared to its rejections. 2010 was not the year to be a consumer taking a loan from lending club. looking towards the present we can see that the fluctuations have gone down a lot and averages roughly around 80% change in loans accepted vs loans rejected. However, arguements can be made that this is not entirely due to lending club being unsuccessful as before, as it could also be attributed to a healthy economy. During the recession many people would have needed small loans to pay bills or make purchases. But, since around 2012, unemployment rate has been steadly declining meaning more people can make purchases or have savings developed leading to less loans being applied for.

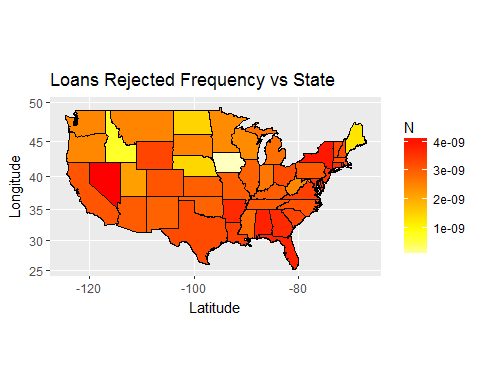
### Relative Frequency of Loans vs. Lower Fourty-Eight States:

Normalized over two values to account for # of loans requested and State population Size.

LAFmap <- LAFmap[order(LAFmap$order),]  
ggplot(data = LAFmap, aes(x=long, y=lat, group = group)) +  
 geom\_polygon(aes(fill = N )) + geom\_path() +  
 scale\_fill\_gradientn(colours=rev(heat.colors(10)),na.value="grey10") + coord\_map()+  
 xlab('Latitude') + ylab('Longitude')+  
 ggtitle('Loans Accepted Frequency vs State')



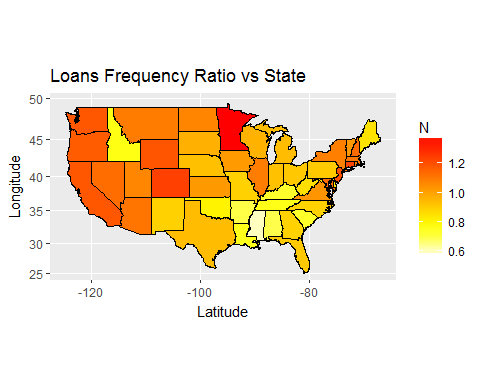
LRFmap <- LRFmap[order(LRFmap$order),]  
ggplot(data = LRFmap, aes(x=long, y=lat, group = group)) +  
 geom\_polygon(aes(fill = N)) + geom\_path() +  
 scale\_fill\_gradientn(colours=rev(heat.colors(10)),na.value="grey10") + coord\_map()+  
 xlab('Latitude') + ylab('Longitude')+  
 ggtitle('Loans Rejected Frequency vs State')



This final plot out of the three below is a ratio of number of loans accepted vs number of loans rejected.

We can see that lending club tends to reject more loans from south eastern united states while accepting more loans towards the north east and west coast areas.

LDFmap <- LDFmap[order(LDFmap$order),]  
ggplot(data = LDFmap, aes(x=long, y=lat, group = group)) +  
 geom\_polygon(aes(fill = N)) + geom\_path() +  
 scale\_fill\_gradientn(colours=rev(heat.colors(10)),na.value="grey10") + coord\_map()+  
 xlab('Latitude') + ylab('Longitude')+  
 ggtitle('Loans Frequency Ratio vs State')

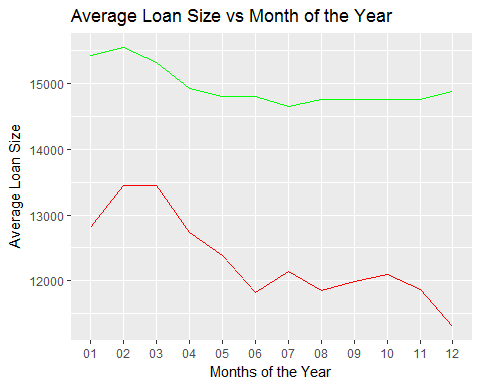


### Average Loan Size vs. Month of the Year:

This group of plots is created to show the average Loan sizes accepted and rejected Based on the month of the year.

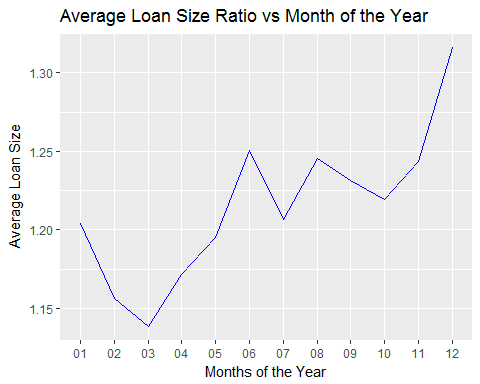
Across the board we can see larger loans tend to be accepted while smaller loans tend to be rejected. However, if a consumer is interested in taking out a smaller loan the best time to do so is towards the end of the year, as lending clubs average rejected loan size decreasese continually as it approaches towards december

ggplot()+   
geom\_line(data=LALS\_MoTY, aes(x=AppMo, y=V1, group=1, color='Accepted Loan'), color='green')+  
geom\_line(data=LRLS\_MoTY, aes(x=AppMo, y=V1, group=1, color='Rejected Loan'), color='red')+  
 xlab('Months of the Year') + ylab('Average Loan Size')+  
 ggtitle('Average Loan Size vs Month of the Year')



This plot below is the ratio between loan size accepted and loan size rejected per month. This clearly shows the trend we could see in our first charts, that as we approach the end of the year the average rejected loan size decreases faster than the changes in average accepted loan size.

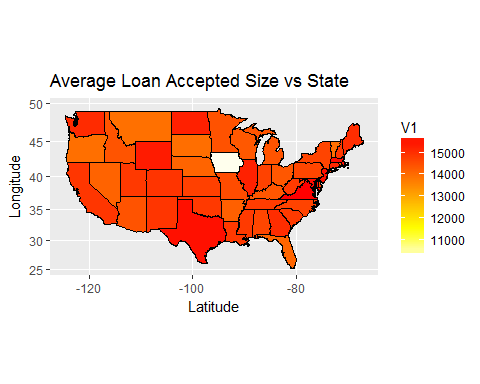
ggplot()+   
geom\_line(data=LDLS\_MoTY, aes(x=AppMo, y=V1, group=1, color='Accepted Loan'), color='blue')+  
 xlab('Months of the Year') + ylab('Average Loan Size')+  
 ggtitle('Average Loan Size Ratio vs Month of the Year')



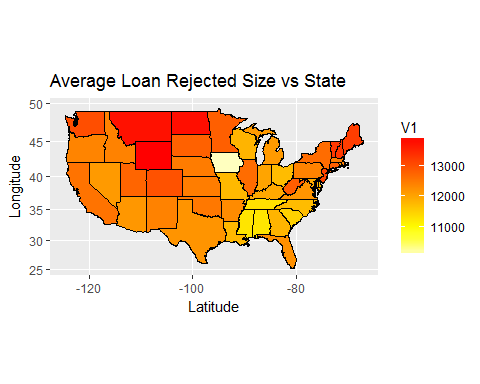
### Loan Size vs. Lower Fourty-Eight States:

Below is a set of heat map plots over the lower fourty-eight united states showing the average loan sizes of each state accepted and rejected. The interesting part of these plots is Iowa, Iowa is the only state with an average accepted loan size lower than its average rejected loan size, which is why our first chart looks so dark.

LALSmap <- LALSmap[order(LALSmap$order),]  
ggplot(data = LALSmap, aes(x=long, y=lat, group = group)) +  
 geom\_polygon(aes(fill = V1)) + geom\_path() +  
 scale\_fill\_gradientn(colours=rev(heat.colors(30)),na.value="grey40") + coord\_map()+  
 xlab('Latitude') + ylab('Longitude')+  
 ggtitle('Average Loan Accepted Size vs State')

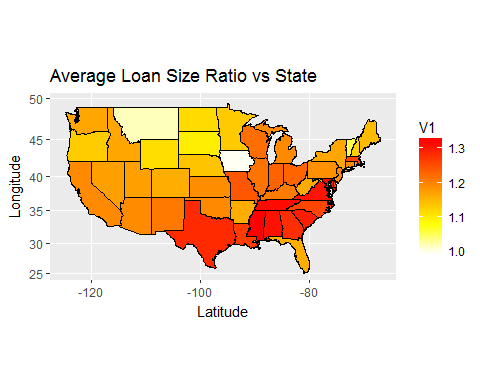


LRLSmap <- LRLSmap[order(LRLSmap$order),]  
ggplot(data = LRLSmap, aes(x=long, y=lat, group = group)) +  
 geom\_polygon(aes(fill = V1)) + geom\_path() +  
 scale\_fill\_gradientn(colours=rev(heat.colors(10)),na.value="grey10") + coord\_map()+  
 xlab('Latitude') + ylab('Longitude')+  
 ggtitle('Average Loan Rejected Size vs State')



The plot below is the best one to examine this peice of data. If you're taking out a small loan the best place to go is Iowa. However, if you are taking out a large loan we can see the best place to do so is the south east and Texas, with florida being the only exception.

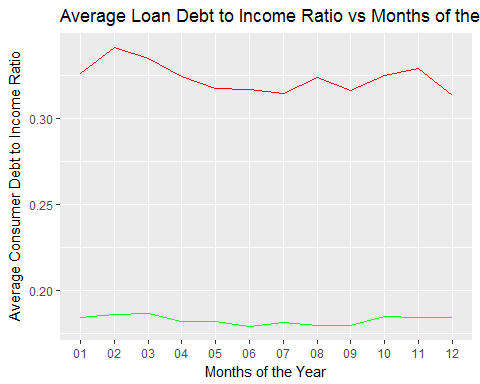
LDLSmap <- LDLSmap[order(LDLSmap$order),]  
ggplot(data = LDLSmap, aes(x=long, y=lat, group = group)) +  
 geom\_polygon(aes(fill = V1)) + geom\_path() +  
 scale\_fill\_gradientn(colours=rev(heat.colors(48)),na.value="grey10") + coord\_map()+  
 xlab('Latitude') + ylab('Longitude')+  
 ggtitle('Average Loan Size Ratio vs State')



### Average Consumer Debt to Income Ratio vs Month of the Year:

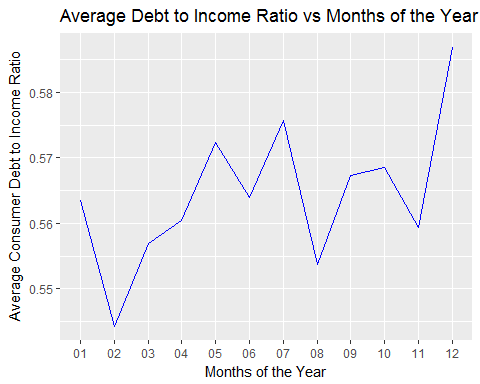
These group of plots show the average accepted and rejected loan's debt to income ratio at the time the consumer applied for it. we can see that unlike loan size higher debt to income ratios are rejected while lower ones are accepted.

ggplot()+   
geom\_line(data=LADTI\_MoTY, aes(x = AppMo, y=V1, group=1, color='Accepted Loan'), color='green') +  
geom\_line(data=LRDTI\_MoTY, aes(x=AppMo, y=V1, group=1, color='Rejected Loan'), color='red') +  
 xlab('Months of the Year') + ylab('Average Consumer Debt to Income Ratio')+  
 ggtitle('Average Loan Debt to Income Ratio vs Months of the Year')



An interesting trend we can spot here is again as we approach december the ratio between accepted loan debt to income ratio and rejected loan debt to income ratio goes up dramatically. This shows a clear trend that as we approach the later months of the year lending club relaxes their requirments for taking out a loan.

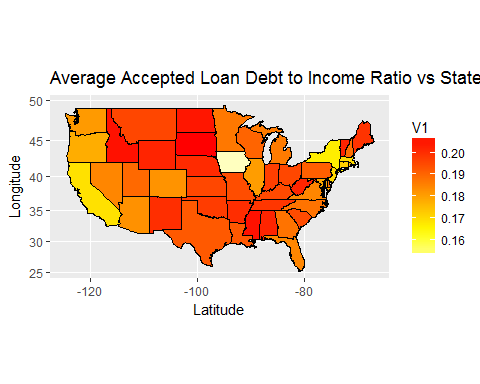
ggplot()+   
geom\_line(data=LDDTI\_MoTY, aes(x=AppMo, y=V1, group=1, color='Accepted Loan'), color='blue')+  
 xlab('Months of the Year') + ylab('Average Consumer Debt to Income Ratio')+  
 ggtitle('Average Debt to Income Ratio vs Months of the Year ')



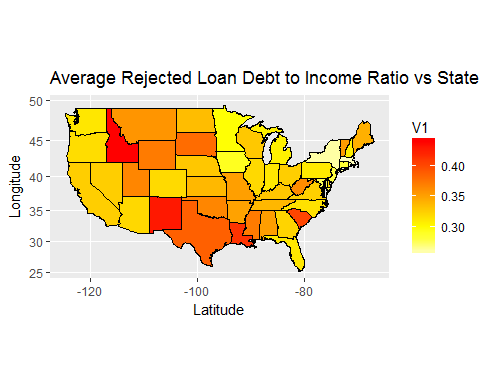
### Average Consumer Debt to Income Ratio vs Lower Fourty-Eight States:

Accepted and Rejected loans debt to income ratio vs lower 48 states. For consumers states with the easiest loans to get would be ones that have higher debt to income ratios on accepted loans and higher debt to income ratios on rejected loans. Below we can see the areas where this occurs is mostly in the mid-western united states.

LADTImap <- LADTImap[order(LADTImap$order),]  
ggplot(data = LADTImap, aes(x=long, y=lat, group = group)) +  
 geom\_polygon(aes(fill = V1)) + geom\_path() +  
 scale\_fill\_gradientn(colours=rev(heat.colors(10)),na.value="grey10") + coord\_map()+  
 xlab('Latitude') + ylab('Longitude')+  
 ggtitle('Average Accepted Loan Debt to Income Ratio vs State')

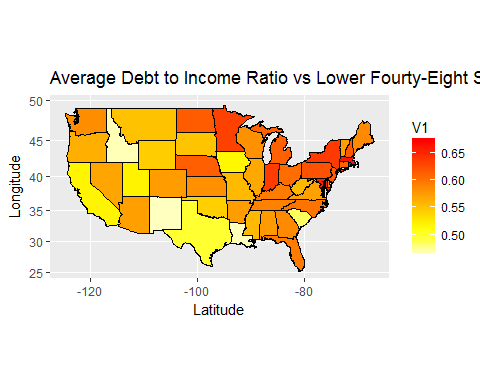


LRDTImap <- LRDTImap[order(LRDTImap$order),]  
ggplot(data = LRDTImap, aes(x=long, y=lat, group = group)) +  
 geom\_polygon(aes(fill = V1)) + geom\_path() +  
 scale\_fill\_gradientn(colours=rev(heat.colors(10)),na.value="grey10") + coord\_map()+  
 xlab('Latitude') + ylab('Longitude')+  
 ggtitle('Average Rejected Loan Debt to Income Ratio vs State')



For our ratio plot, a larger ratio between loans accepted and loans rejected is ideal for consumers since it implies that even with an average loan accepted DTI, if the state average for rejection is higher that consumers DTI could still be adequate.

LDDTImap <- LDDTImap[order(LDDTImap$order),]  
ggplot(data = LDDTImap, aes(x=long, y=lat, group = group)) +  
 geom\_polygon(aes(fill = V1)) + geom\_path() +  
 scale\_fill\_gradientn(colours=rev(heat.colors(10)),na.value="grey10") + coord\_map()+  
 xlab('Latitude') + ylab('Longitude')+  
 ggtitle('Average Debt to Income Ratio vs Lower Fourty-Eight States')

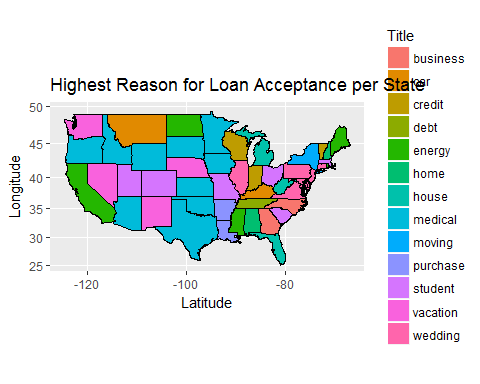


### Loan Title Frequency vs Lower Fourty-Eight States:

Frequency of loan purpose vs lower 48 states. In the last two plots below we've taken the frequency counts of all the accepted loans titles and all the rejected loans titles and checked which state had the highest of each.

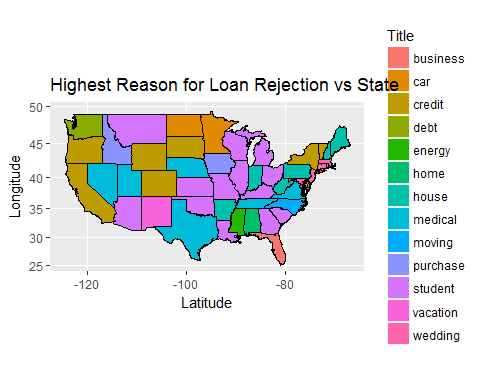
For accepted loans we can see that the most frequently occuring reason is medical. Its good to know that Lending Club is helping people out in their times of need and as a consumer it must feel reassuring to hear.

LATmap <- LATmap[order(LATmap$order),]  
ggplot(data = LATmap, aes(x=long, y=lat, group = group)) +  
 geom\_polygon(aes(fill = Title)) + geom\_path() + coord\_map()+  
 xlab('Latitude') + ylab('Longitude')+  
 ggtitle('Highest Reason for Loan Acceptance per State')



Looking at rejected loans, they're all over the place but it seems that student loans are the most commonly rejected reason. However, student loans outside of the goverment are very risky for investors since unlike with material objects or medical bills student loans don't guarantee that person will have a job or be able to pay these bills back. Further with government student loans there is even less risk since wages can be garnished if you declare bankruptcy, but the same is not true for the private sector.

LRTmap <- LRTmap[order(LRTmap$order),]  
ggplot(data = LRTmap, aes(x=long, y=lat, group = group)) +  
 geom\_polygon(aes(fill = Title)) + geom\_path() + coord\_map()+  
 xlab('Latitude') + ylab('Longitude')+  
 ggtitle('Highest Reason for Loan Rejection vs State')



### Conclusions:

After examining through all of our data from above, I believe we can come to several interesting conclusions. By examining the average consumers debt to income ratio and loan sizes vs months of the year, we can see a trend as the end of the year approaches where the number of loans rejected drops dramatically. From this we can assume that as the years end approaches Lending Club sales people are trying to meet end of the year quota so less satisfactory loans will be accepted to make ends meet. Another assumption we can make is during the last few months of the year the holiday season is here. Many people are Christmas shopping and therefore need smaller loans to make purchases for friends and family. While some people would still get rejected the number of people who are looking for a small loan go from people who need a payday loan to all sorts of folks.

As for locational trends, We can notice two additional trends. As we look towards the south eastern part of the united states the ratio between accepted loan size and rejected loan size is the largest in the country, and concentrated to mostly this area. The other trend we see comes from consumer debt to income ratio, in the mid west we can see has the largest differences between accepted loans debt to income ratio and rejected loans debt to income ratio. Both of these can be understood by examining the local populations closer. For the south-eastern united states we know on average have lower wages and would tend to request smaller loan sizes, further if a higher percentage of the population tends to live closer to the poverty line, they would have other outstanding reasons as to why even though they have some of the smallest loan rejected sizes they have some of the highest percentage of loan rejections vs state population. Looking at the mid-west with and its trend in consumer debt to income ratio, its the oppisite. while there is still higher instances of people living closer to the poverty line out in this region there are still very few people living in these areas vs the larger coastal cities. For Lending Club to get its business out there and have consumers hear about them, they would need to lower the requirements for a loan to get accepted so they can spread their business through word of mouth.