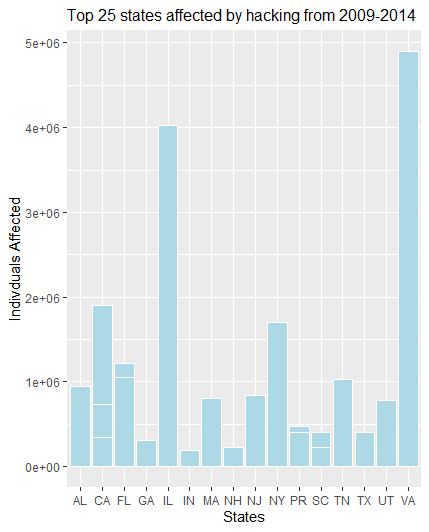
CIS 435 Project Question 3 of group 6

The ways I tried and analyzed the dataset is to construct a visual bar graph and base the top 25 states affected by hacking from 2009-2015 and in each bar for each state each hacking attempted is grouped for each individual incident. I then clean the data further by taking care of all things related to subcategories for theft improper disposal and so and changed the categories to just those types. Finally, I used the summary on the data for the type of breach to give a list of categories and how many of each of those cases occur in the time period of 2009 to 2014. The things I will still have to do is to figuring out what the trend is for the entire dataset and not just the top 25 states and to see if there are different patterns in the states with lower affected individuals.



Hacking/IT Incident

76

Improper Disposal

39

Loss

98

Other

91

Theft

571

Unauthorized Access/Disclosure

168

Unknown

12

top25 <-Projectdata[Projectdata$Individuals\_Affected >= 185000,]

ggplot(top25, aes(x=top25$State,y=top25$Individuals\_Affected)) + geom\_bar(stat = "identity",fill="lightblue",colour ="white",position = "dodge") + xlab("States") + ylab("Indivduals Affected") + ggtitle("Top 25 states affected by hacking from 2009-2014") + theme(plot.title = element\_text(size = 12))

Projectdata$Type\_of\_Breach[Projectdata$Type\_of\_Breach == "Loss, Unauthorized Access/Disclosure"] <- "Loss"

Projectdata$Type\_of\_Breach[Projectdata$Type\_of\_Breach == "Loss, Improper Disposal"] <- "Loss

Projectdata$Type\_of\_Breach[Projectdata$Type\_of\_Breach %in% c("Theft, Hacking/IT Incident","Theft, Improper Disposal, Unauthorized Access/Disclosure","Theft, Loss","Theft, Loss, Improper Disposal","Theft, Loss, Other","Theft, Loss, Unauthorized Access/Disclosure, Unknown","Theft, Other","Theft, Unauthorized Access/Disclosure","Theft, Unauthorized Access/Disclosure, Hacking/IT Incident"," Theft, Unauthorized Access/Disclosure, Other")] <- "Theft"

Projectdata$Type\_of\_Breach[Projectdata$Type\_of\_Breach == "Theft, Unauthorized Access/Disclosure, Other"] <- "Theft"

Projectdata$Type\_of\_Breach[Projectdata$Type\_of\_Breach %in% c("Unauthorized Access/Disclosure, Hacking/IT Incident","Unauthorized Access/Disclosure, Hacking/IT Incident, Other","Unauthorized Access/Disclosure, Other")] <- "Unauthorized Access/Disclosure"

Projectdata$Type\_of\_Breach[Projectdata$Type\_of\_Breach == "Unknown, Other"] <- "Unknown"

Projectdata$Type\_of\_Breach[Projectdata$Type\_of\_Breach == "Hacking/IT Incident, Other"] <- "Hacking/IT Incident"

"Improper Disposal, Unauthorized Access/Disclosure"] <- "Improper Disposal"

summary(Projectdata$Type\_of\_Breach)