R Plot Draft

Our research questions ask the relationship between obesity and potentially dependent variables such as income, education level, and state of residence, in order to inform priorities for public health programs aimed at reducing obesity. For example, how do obesity rates correlate with income? Does obesity increase as income level decreases? The dataset divided the income levels as: Less than $15,000; $15,000 - $24,999; $25,000 - $34,999; $35,000 - $49,999; $50,000 - $74,999 and $75,000 or greater. We expected to find that as income level decreases, the rate of obesity increases. By plotting our income and obesity data, we can show the inverse relationship between income and obesity. The various types of plots below each have their own advantages for visualizing the data and relationships.

We adjusted our plots to look across all four years of the survey data, 2011 - 2015, and included several plots showing the overall increase in obesity rates during this time period.

To create those plots, we installed the following packages:

install.packages("xkcd",dependencies = TRUE)

readr, ggplot2, plyr, extrafont, xkcd.

library(ggplot2)

library(extrafont)

library(xkcd)

loadfonts(device="win")

font\_import(pattern="[X/x]kcd")

Two objects used in the plots are Obesity Average by Income and Obesity Average by Location

obesity\_average\_by\_income$Income[obesity\_average\_by\_income$Income=="$15,000 or less"] <- "$0 - $15,000"

average\_by\_location\_reorder=obesity\_average\_by\_location[order(obesity\_average\_by\_location$Data\_Value),]

average\_by\_location\_reorder$LocationDesc=factor(average\_by\_location\_reorder$LocationDesc,levels=average\_by\_location\_reorder$LocationDesc)

View(average\_by\_location\_reorder)

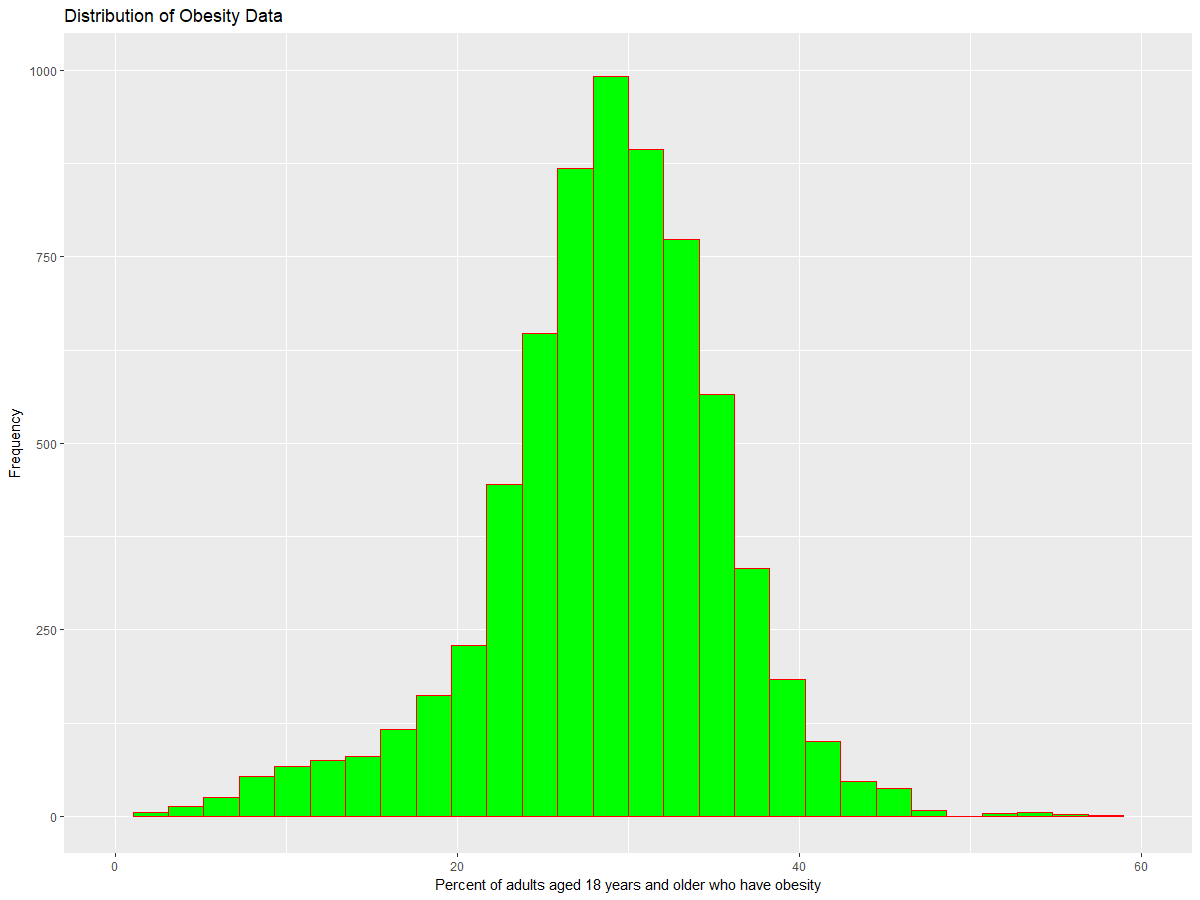
Obesity Distribution Plot & One Variable Density Plot

To start we created a histogram and a one variable density plot, both of obesity rates across all potentially dependent variables, to get a visual overview of the data.

#Histogram for Obesity Distribution

ggplot(percentobesity,aes(x=Data\_Value))+geom\_histogram(color="red",fill="green")+ggtitle("Distribution of Obesity Data")+

ylab("Frequency")+xlab("Percent of adults aged 18 years and older who have obesity")+xlim(0,60)+ylim(0,1000)

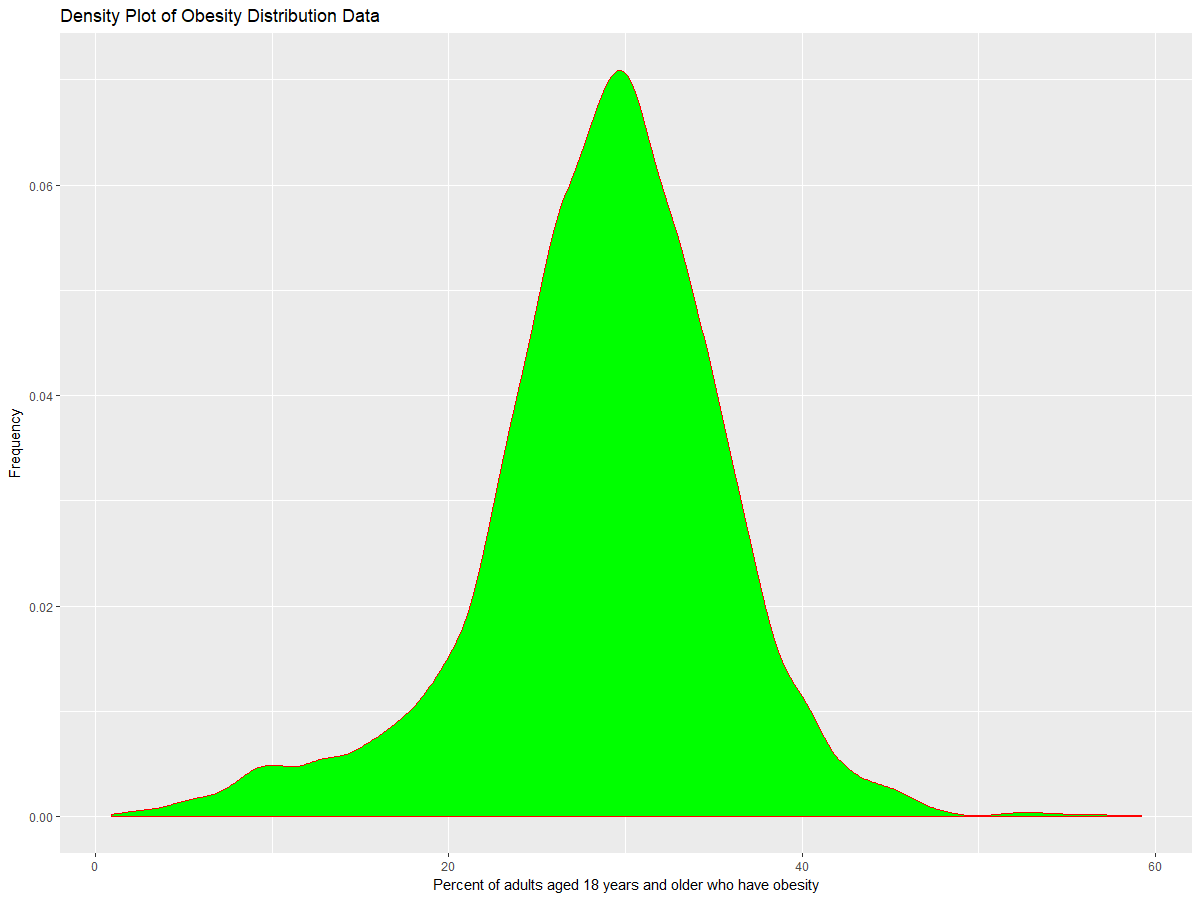


#One Variable Density Plot

ggplot(percentobesity,aes(x=Data\_Value))+geom\_density(color="red",fill="green")+

ggtitle("Density Plot of Obesity Distribution Data")+

ylab("Frequency")+xlab("Percent of adults aged 18 years and older who have obesity")

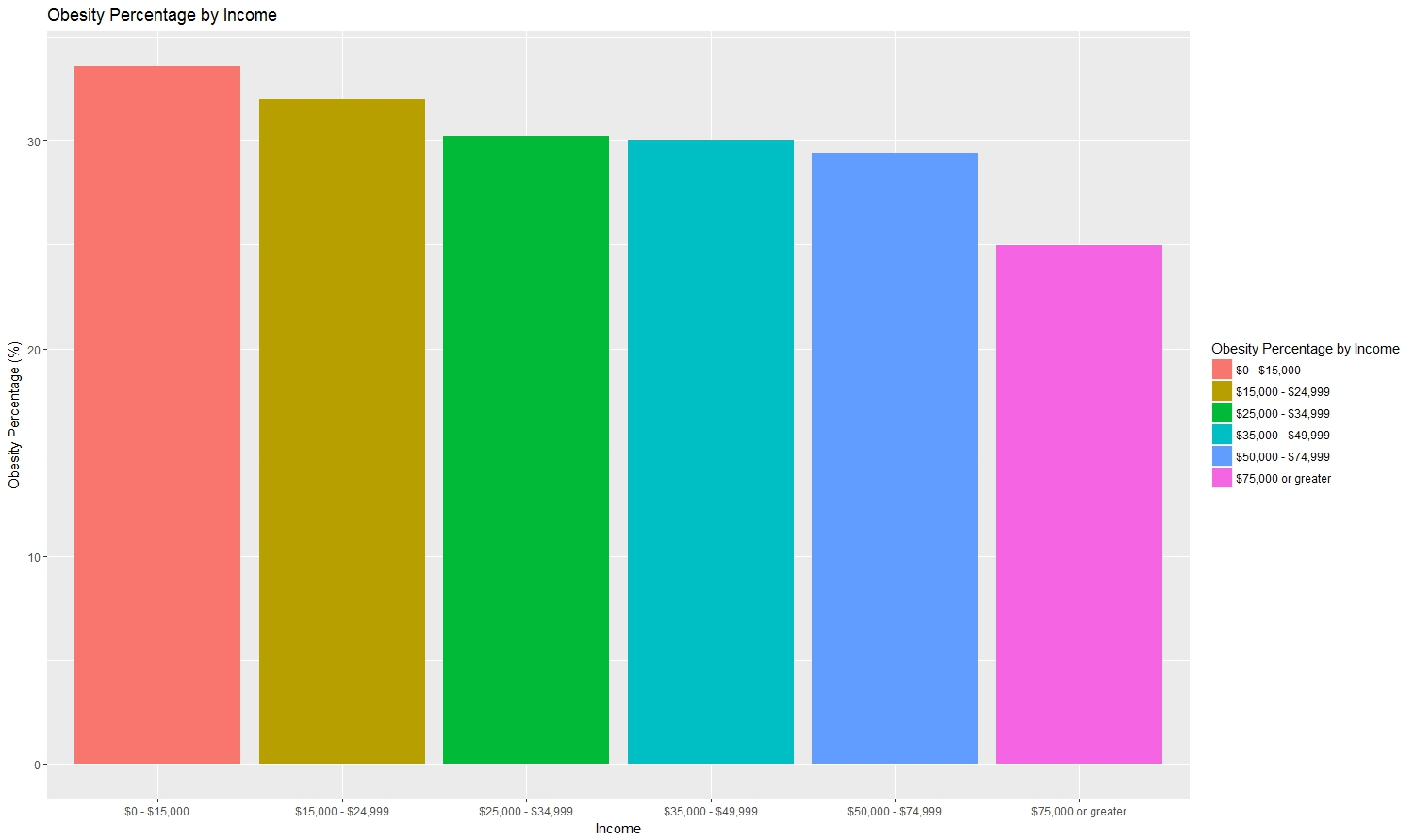


Bar Plot for Obesity and Income Level

The bar plot below shows that the rate of obesity decreases as the income level increases. Confirming this relationship can be helpful to take preventive steps to decrease obesity in low income level households.

incg <- ggplot(obesity\_average\_by\_income, aes(x=Income, y=Data\_Value, , fill=Income)) + geom\_col()

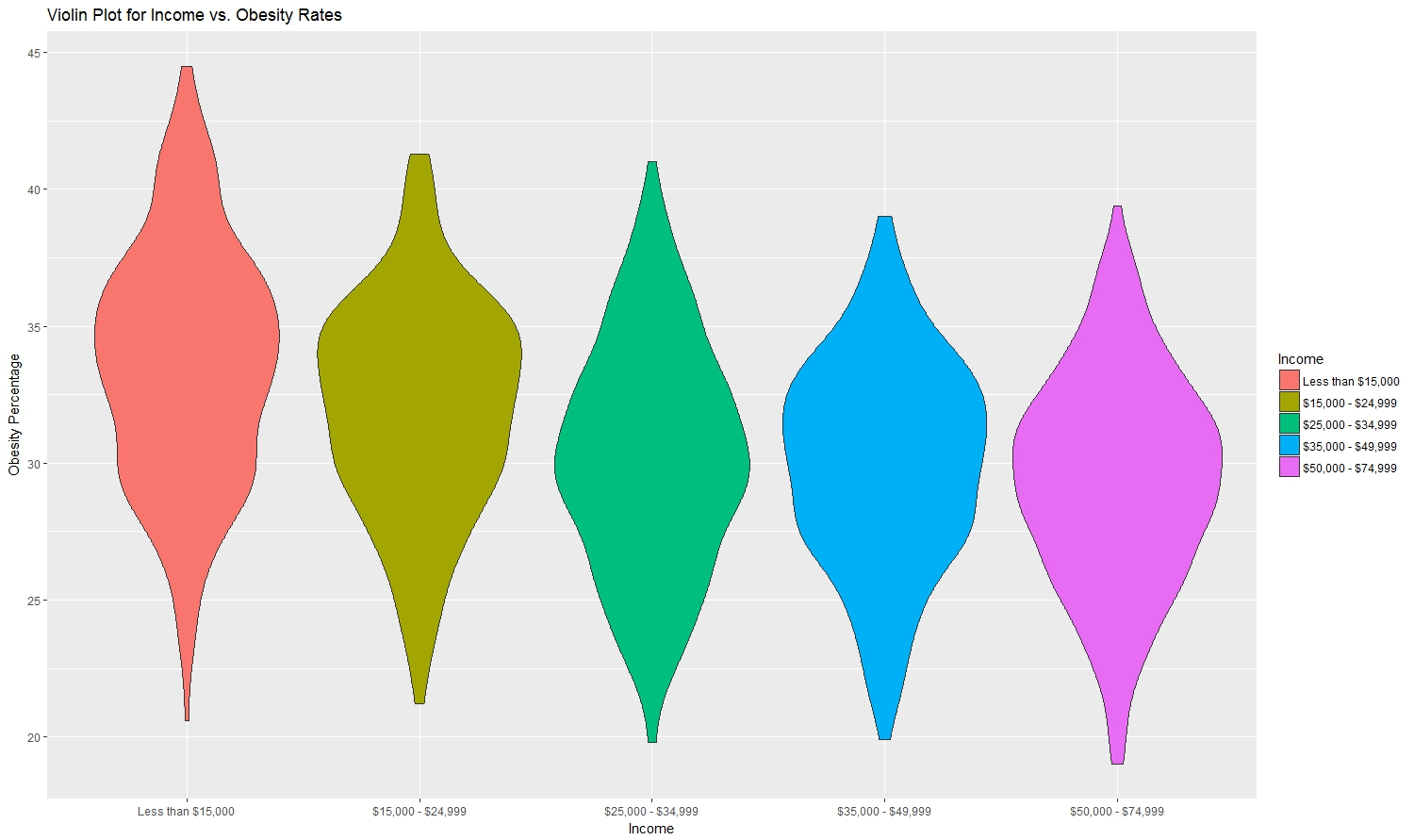
incg + labs(title='Obesity Percentage by Income', y='Obesity Percentage (%)',x='Income') + guides(fill=guide\_legend(title="Obesity Percentage by Income"), colour=guide\_legend(reverse=TRUE))



Violin Plot for Average Obesity Rates by Income

Violin plots look similar to box-plots, but differ slightly in that they show the distribution of each category. Like box plots, violin plots are used to represent comparisons of a variable distribution across different categories. Since we wanted to compare rates of obesity across different income levels, we used a violin plot here.

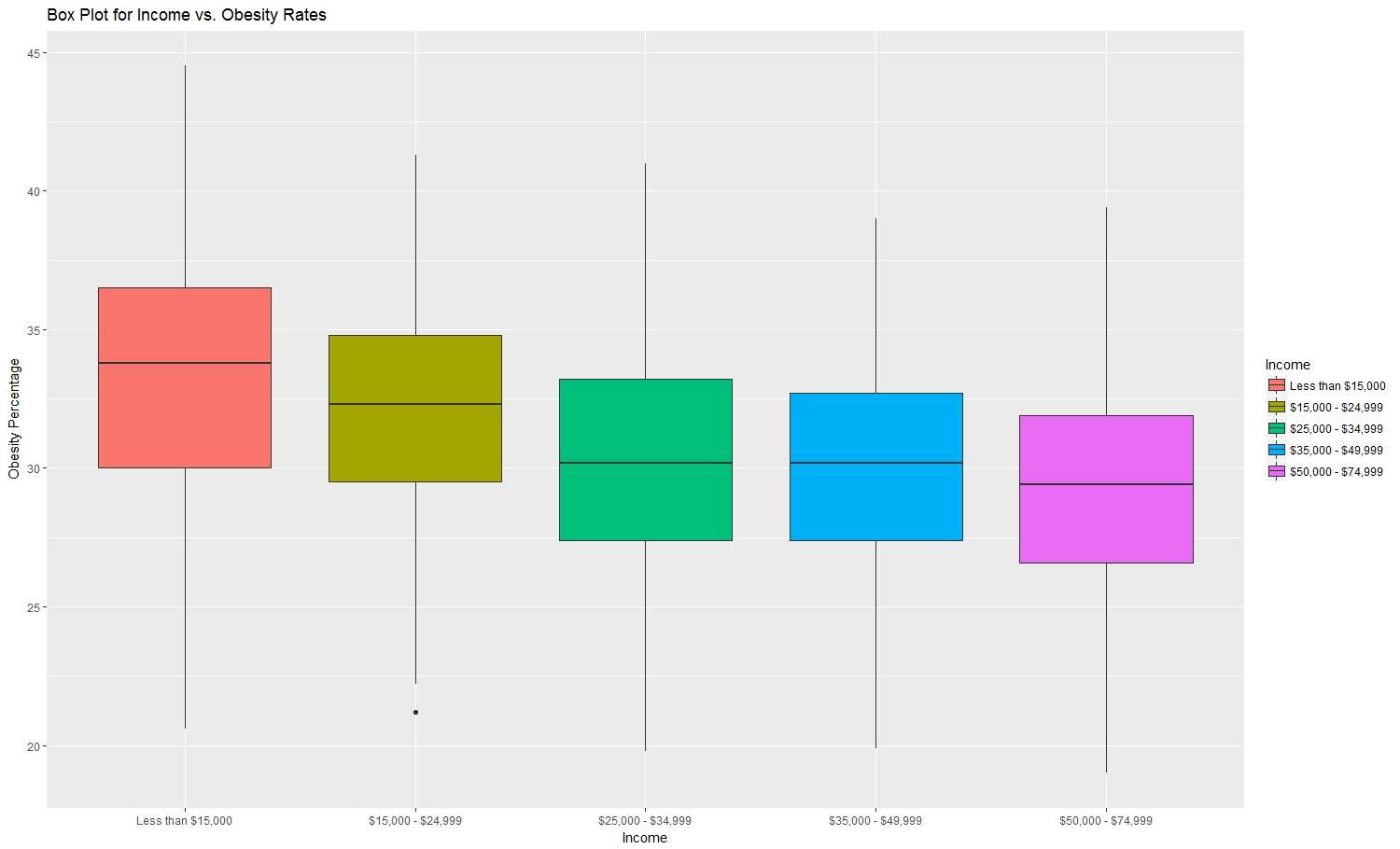
ggplot(data=subset(obesity, !is.na(obesity$Income)),aes(x=Income,y=Data\_Value,fill=Income)) + geom\_violin() + ylab("Obesity Percentage") + ggtitle("Violin Plot for Income vs. Obesity Rates")



Box Plot for Average Obesity Rates by Income

The advantage of using a box-plot is that it not only helps to compare obesity across income categories, but also provides extra details such as mean, maximum, minimum, and outliers.

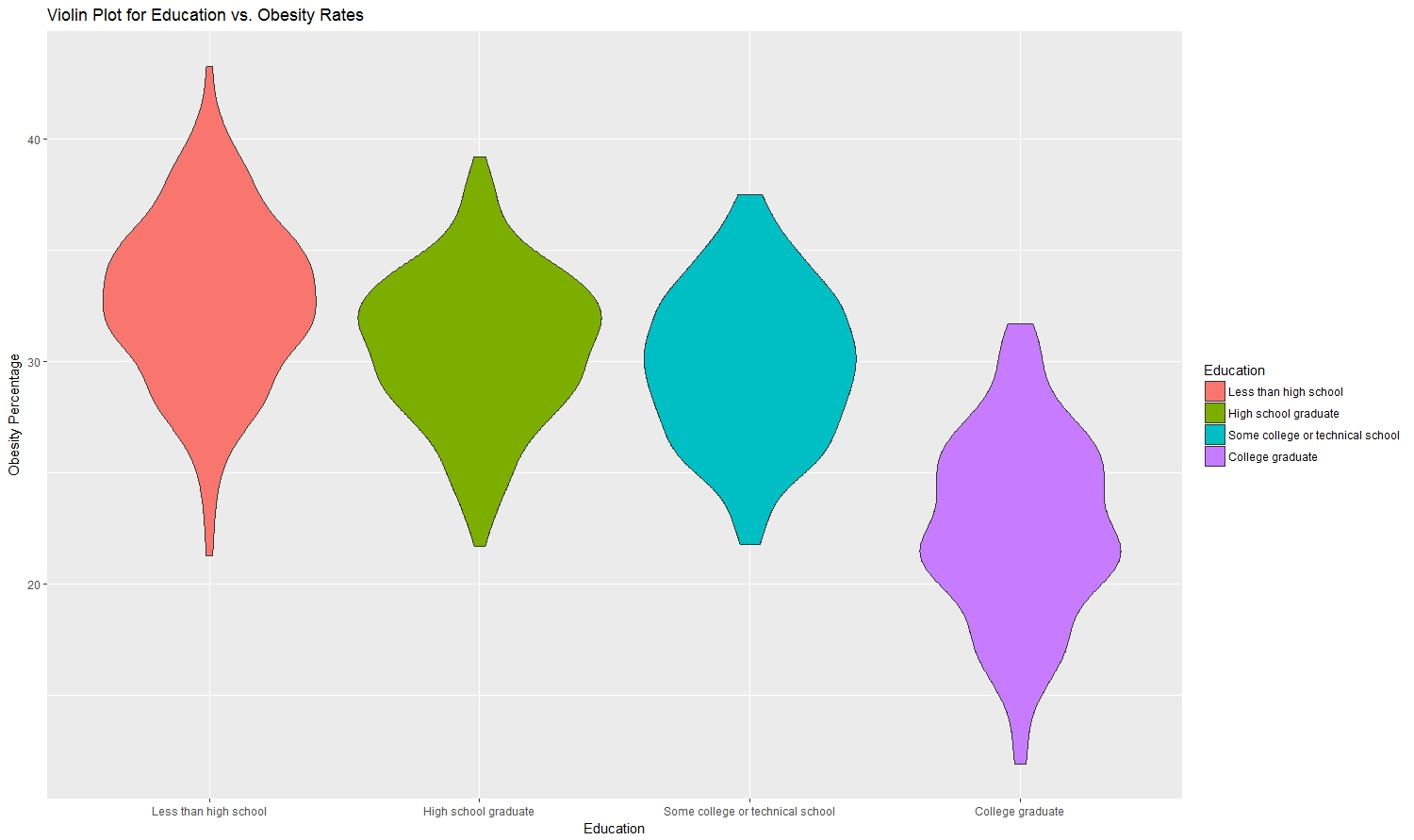
ggplot(data=subset(obesity, !is.na(obesity$Income)),aes(x=Income,y=Data\_Value,fill=Income))+geom\_boxplot()+ylab("Obesity Percentage")+ggtitle("Box Plot for Income vs. Obesity Rates")



Violin Plot for Average Obesity Rates by Education

Income is closely related with other demographic factors, like education level. Since the relationship between obesity and income was consistent with our theory, we wanted to study related factors more closely, to better inform education and prevention programs. For this purpose, we picked the education variable and used a violin plot to see if there is an inverse relationship similar to income and obesity.

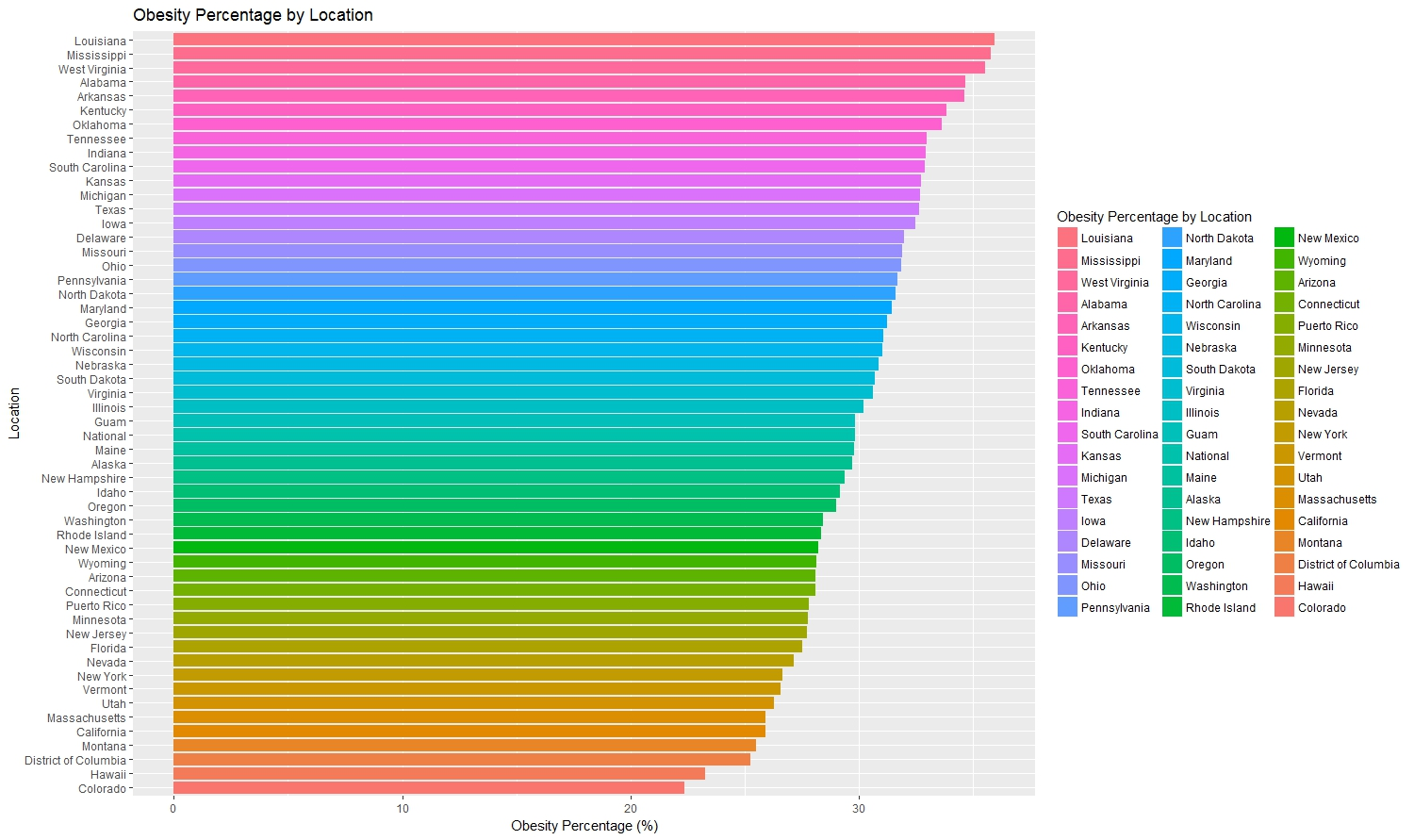
ggplot(data=subset(obesity, !is.na(obesity$Education)),aes(x=Education,y=Data\_Value,fill=Education)) + geom\_violin() + ylab("Obesity Percentage") + ggtitle("Violin Plot for Education vs. Obesity Rates")



R Plot for Average Obesity Rates by Location (States, Territories)

This bar chart shows the relationship between the US locations (states and territories) and obesity percentage distribution across United States. Decision makers can use this bar chart to prioritize the allocation of resources and programming aimed at healthier habits and reducing obesity.

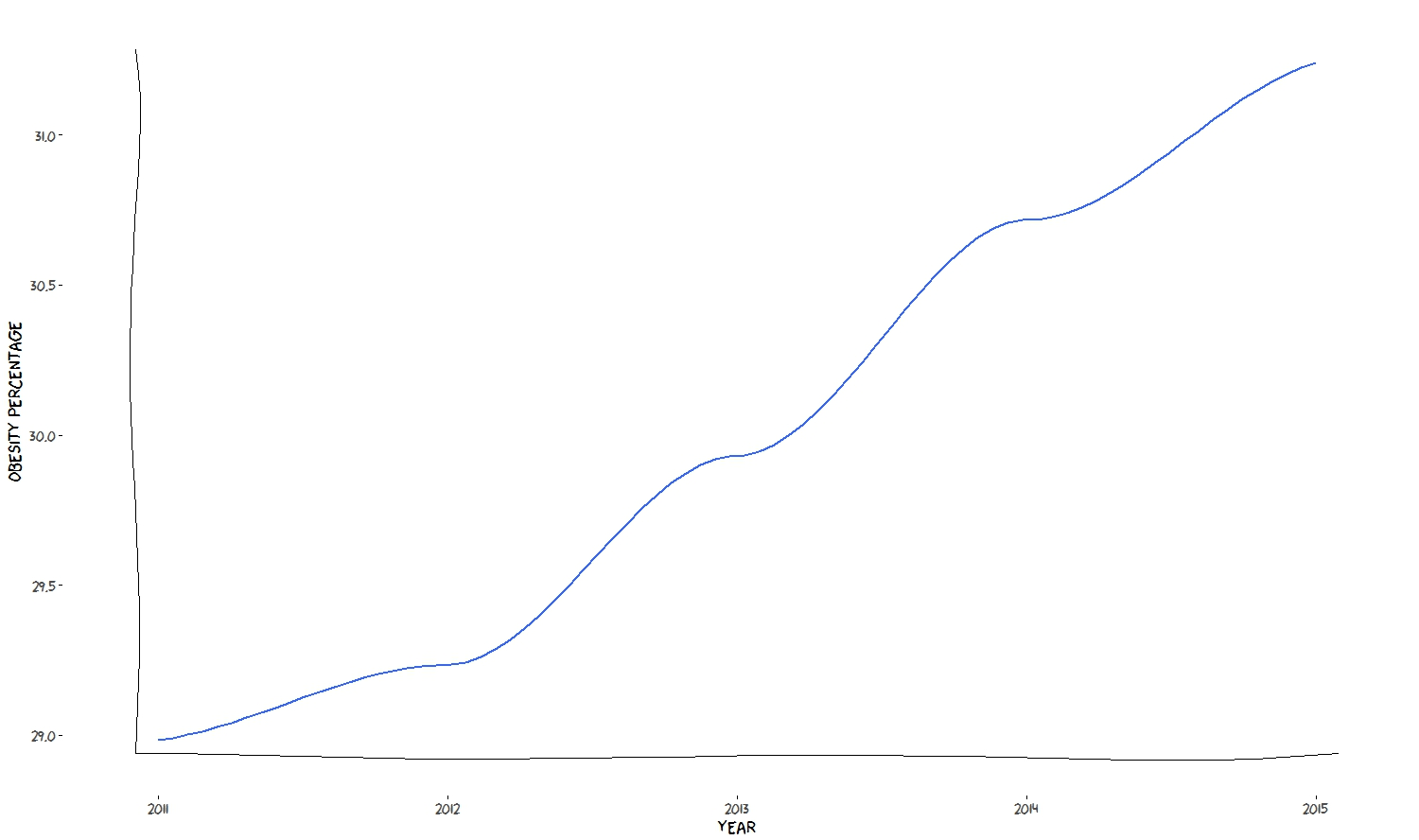
ggplot(average\_by\_location\_reorder,aes(x=factor(LocationDesc),y=Data\_Value, fill=LocationDesc)) + geom\_bar(stat='identity') + coord\_flip() + labs(title = "Obesity Percentage by Location", y='Obesity Percentage (%)',x='Location') + guides(fill=guide\_legend(title="Obesity Percentage by Location", reverse=TRUE), colour=guide\_legend(reverse=TRUE))

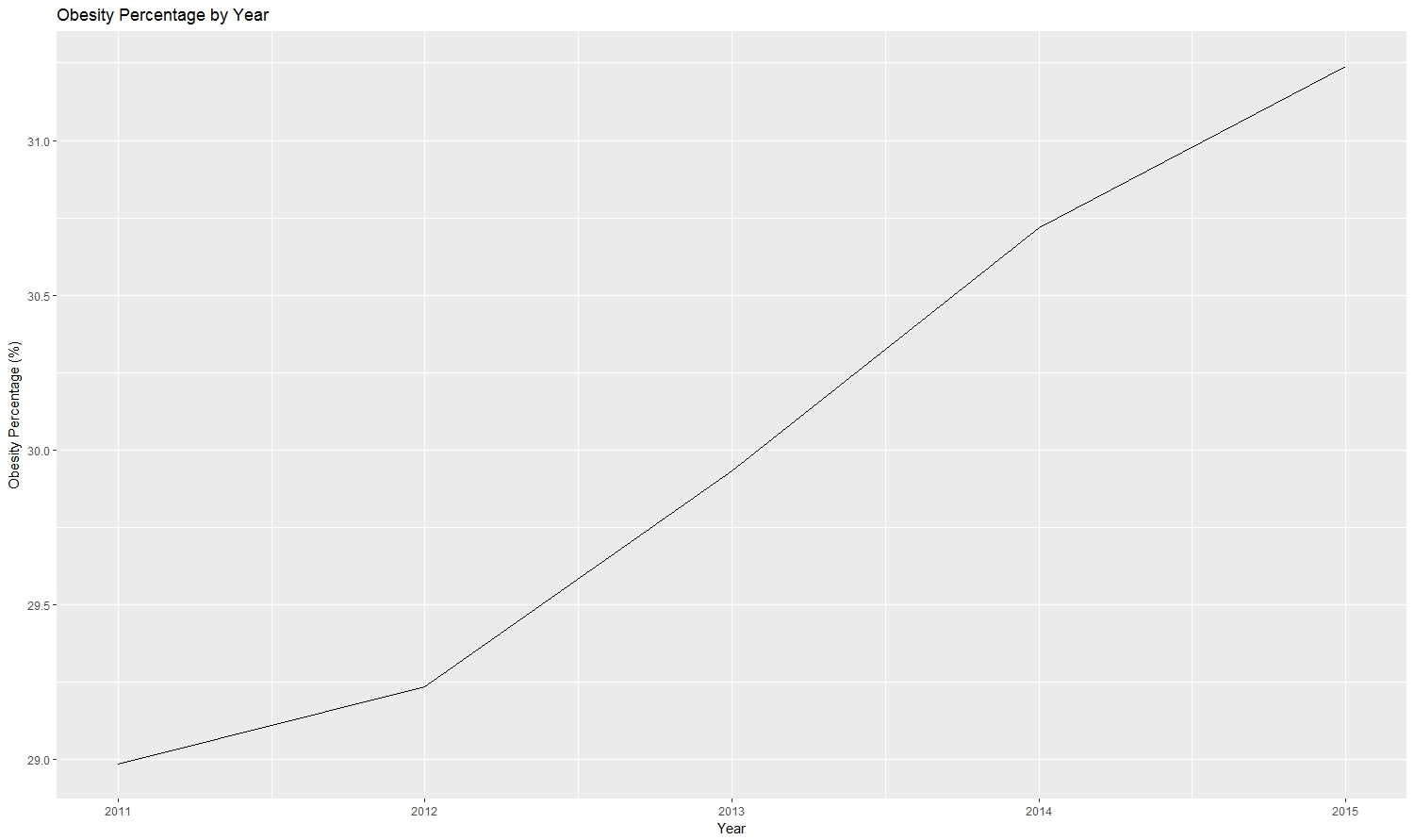


R Plots for Increase in Obesity Rates from 2011 – 2015

In the following three plots, we plot the increase in obesity rates from 2011 to 2015. For the first two the y axis does not start at zero, so the plot clearly shows an increase but the scale of the increase is exaggerated. The first plot uses the xkcd package.

ggplot() + geom\_smooth(mapping=aes(x=Year, y =Data\_Value), data =obesity\_average\_by\_year, method="loess") + xkcdaxis(xrange,yrange) + ylab("Obesity Percentage") + theme(text=element\_text(size=16, family="xkcd"))

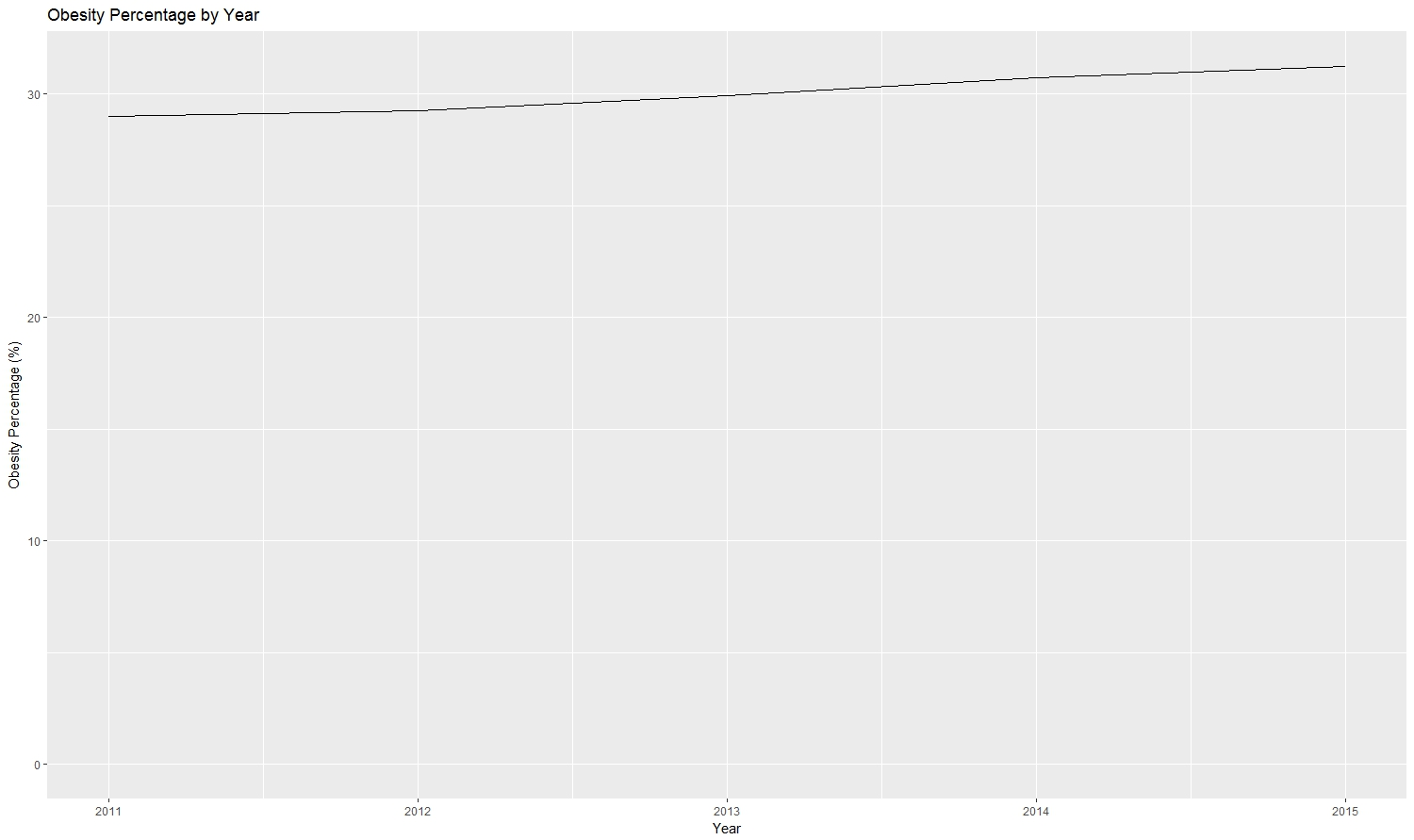




In the plot below we adjusted the y axis to start at zero. It is not a particularly attractive nor powerful visualization, but the increase is represented to an accurate scale.

yearg <- ggplot(obesity\_average\_by\_year, aes(x=Year, y=Data\_Value)) + geom\_line() + labs(title='Obesity Percentage by Year', y='Obesity Percentage (%)',x='Year')

yearg + expand\_limits(y = 0)



Word Count: 762