Data Analysis of the Money Expenditure on Gambling in Australia

Shehab Shahin

University of Sydney | 26/10/2018

Table of Contents

# Executive Summary

This report aims to obtain a better understanding of how would the money spent on gambling increase or decrease as the gambling category differ. The report also tries to observe what state losses the largest amount of money to gambling companies and firms. This report utilize the data set “Expenditure on Gambling” sourced from Queensland Government Statistician’s Office to try to observe and study the trends in data with the aid of various charts and plots.

Section one of the report focuses on the initial data analysis, in which the data is explained. The explanation of the data includes providing a background to the data set, an evaluation of the source of data and an identification of the limitations surrounding the data set.

Section two of the report focuses on answering two questions in relation to the data. The first question attempts to understand which state produces the highest expenditure values by looking at box plots and a mean bar chart and it was found that the state of NSW had the highest gambling expenditure. The second question revolves around finding the relationship between the category of gambling and the money expenditure (In millions of Dollars). The answer for this question also utilizes box plots to investigate important trends in the data, and a mean bar chart that presents the average money expenditure for each gambling category, also, it was found that the gaming category produced the highest average gambling expenditure when compared to the other two categories.

# Full Report

## Initial Data Analysis (IDA)

Gambling = read.csv("Gamb.csv", header= T)  
  
message("Structure of the data which consists of both quantitative and qualitative variables.")

## Structure of the data which consists of both quantitative and qualitative variables.

str(Gambling)

## 'data.frame': 88 obs. of 8 variables:  
## $ State : chr "ACT" "ACT" "ACT" "ACT" ...  
## $ GamblingCategory: chr " Racing " " Racing " " Racing " " Gaming " ...  
## $ GamblingType : chr " Oncoursebookmaker " " OncourseTotalisator " " TAB " " Casino " ...  
## $ Expenditure : num 0.076 0.185 22.089 16.84 167.454 ...  
## $ X : logi NA NA NA NA NA NA ...  
## $ X.1 : logi NA NA NA NA NA NA ...  
## $ X.2 : logi NA NA NA NA NA NA ...  
## $ X.3 : num NA NA NA NA NA NA NA NA NA NA ...

D = dim(Gambling)  
message("Dimensions of the data are: ", toString(D))

## Dimensions of the data are: 88, 8

names(Gambling)

## [1] "State" "GamblingCategory" "GamblingType" "Expenditure"   
## [5] "X" "X.1" "X.2" "X.3"

head(Gambling)

## State GamblingCategory GamblingType Expenditure X X.1 X.2 X.3  
## 1 ACT Racing Oncoursebookmaker 0.076 NA NA NA NA  
## 2 ACT Racing OncourseTotalisator 0.185 NA NA NA NA  
## 3 ACT Racing TAB 22.089 NA NA NA NA  
## 4 ACT Gaming Casino 16.840 NA NA NA NA  
## 5 ACT Gaming GamingMachines 167.454 NA NA NA NA  
## 6 ACT Gaming InstantLottery 1.991 NA NA NA NA

tail(Gambling)

## State GamblingCategory GamblingType Expenditure X X.1 X.2  
## 83 WA Gaming Lotto 327.771 NA NA NA  
## 84 WA Gaming MinorGaming 23.090 NA NA NA  
## 85 WA Gaming Pools 0.734 NA NA NA  
## 86 WA SportsBetting BookmakerandotherFixedOdds 0.005 NA NA NA  
## 87 WA SportsBetting TabFixedOdds 71.451 NA NA NA  
## 88 WA SportsBetting TabToteOdds 0.330 NA NA NA  
## X.3  
## 83 NA  
## 84 NA  
## 85 NA  
## 86 NA  
## 87 NA  
## 88 NA

### Background to data:

Ethics: the work of the QGSO agency is governed by a number of laws and legislation such as the Statistical Returns Act 1896. This law in addition to other laws and legislation assist the agency in obtaining data easily but simultaneously ensuring that the data are only used for research purposes.  
Source: The source of data was the “Queensland Government Statistician’s Office”. This establishment is part of Queensland Treasury. Therefore, the data provided by this organization is reliable as its based on the research conducted by reputable individuals from a different background. Limitations: Within the data set, there were many missing data that led to a significant change in the overall results and specifically in the expenditure values as more than 10 missing expenditure values were in the data set.

### Classification of variables:

There are 4 variables present within this dataset, state, Gambling Category, Gambling Type and Expenditure (In Millions of Dollars). The expenditure is a continuous numerical data. The state is another variable present within the dataset and this variable is a categorical and ordinal one; composed of more than three categories. Finally, Gambling Category and Gambling Type are two variables that are both categorical, Nominal (non-ordered) and composed of three or more categories.

### Stakeholders:

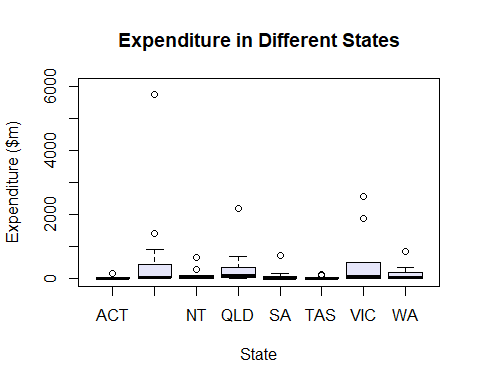
Primary institutions interested:

Government: Use those data to spread awareness of the huge negative impact of gambling.

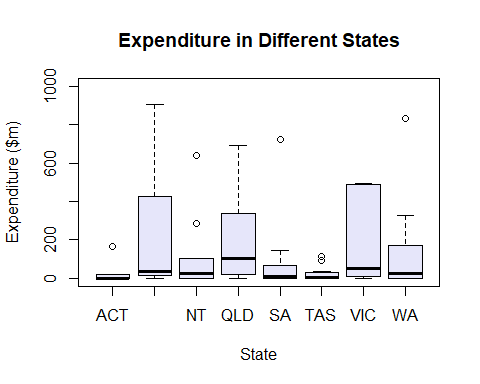
Gambling Companies & Agencies: Observing the trends in money expenditure on different gambling activities which can equip them with sufficient information about the wager’s areas of interest.

## How Does the Money Expenditure Change in Different States ?

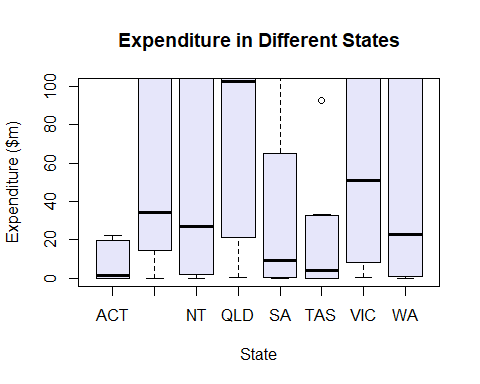
state.expenditure.dist <- split( Gambling$Expenditure, Gambling$State)   
  
boxplot(state.expenditure.dist, col = "lavender",ylim = c(0,6000),ylab = "Expenditure ($m)", xlab = "State", main = "Expenditure in Different States")



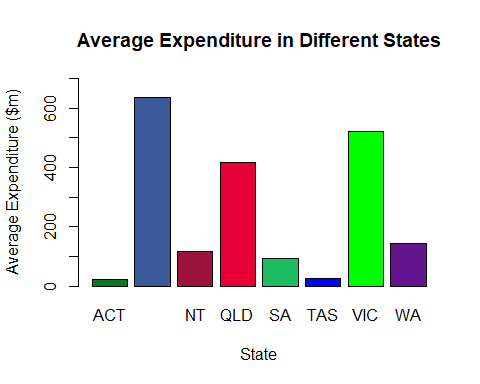
boxplot(state.expenditure.dist, col = "lavender",ylim = c(0,1000), ylab = "Expenditure ($m)", xlab = "State", main = "Expenditure in Different States")



boxplot(state.expenditure.dist, col = "lavender",ylim = c(0,100), ylab = "Expenditure ($m)", xlab = "State", main = "Expenditure in Different States")



state.expenditure.mean <- sapply(state.expenditure.dist, mean)  
  
OB <- c(rgb(12,120,30, maxColorValue = 255), rgb(59,89,152, maxColorValue = 255), rgb(156,20,60, maxColorValue = 255),  
rgb(231,0,56,maxColorValue = 255),  
rgb(30,188,99,maxColorValue = 255),  
rgb(0,0,255,maxColorValue = 255),  
rgb(0,255,0,maxColorValue = 255),  
rgb(100,20,140,maxColorValue = 255))  
  
barplot(state.expenditure.mean, col = OB, ylim = c(0,700), ylab = "Average Expenditure ($m)", xlab = "State", main = "Average Expenditure in Different States")



T = tail(state.expenditure.mean, n=8)  
message ("The Expenditure values From left to Right: ", toString(T ))

## The Expenditure values From left to Right: 23.017, 636.528571428571, 118.1396, 416.088444444444, 93.1496363636364, 25.881375, 523.067272727273, 143.384636363636

### Text and Analysis:

By looking at the box plot, it can be seen that the maximum expenditure value that is not an outliner is present in the state of NSW with a value of $905.783m. the minimum expenditure value in all the states was relatively the same and was close to the value of $0.001m. the median expenditure in each of the states is as follow: QLD: $102.6m, VIC: $51.0m, NSW: $34.5m, NT: $26.9m, WA: $23.1m, SA: $9.4m, TAS: $3.9m, ACT: $1.59m.

The bar chart represents the mean gampling expenditure in each of the Australia states with the highest average expenditure being in the state of NSW with an expenditure value of $636.5m. on the other hand, the lowest average expenditure value was in the state of ACT with an expenditure value of $23.017m.

### Summary:

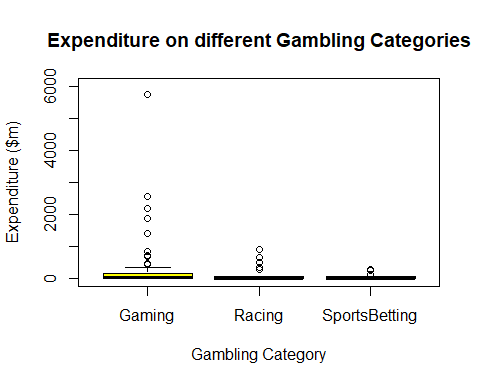
It can be predicted that the states of NSW and VIC would produce the largest expenditure values as those two states are the largest in terms of population. This prediction was verified through the statistical representation of the data using a bar chart which clearly ranks the average expenditure in the different Australian states. The rank was as following (From the Highest to Lowest): NSW, VIC, QLD, WA, NT, SA, TAS, ACT. Furthermore, the boxplot detected a number of important trends that are worth mentioning, for example, when talking about NSW state, it had a median of $34.5m which was calculated using 14 expenditure values from 14 gambling activity, hence 7 out of those 14 expenditure activities produced less than $34.5m while the other 7 generated a very high net income which caused the overall expenditure in NSW to be the highest despite the fact that the state of NSW did not have the highest median. To sum up, the overall high expenditure in NSW was mainly concentrated in 7 gambling activities or places. Moreover, by looking at the median values of the other boxes, it can be concluded that the money expenditure is concentrated on a few or a large number of activities. In conclusion, it was found that the state of NSW had the highest average gambling expenditure which is mainly due to its high population compared to the other Australian states. The final trend was that the higher the population in a state, the higher the gambling expenditure will be.

## What is the Relationship Between the Money Expenditure and the Gambling Category ?

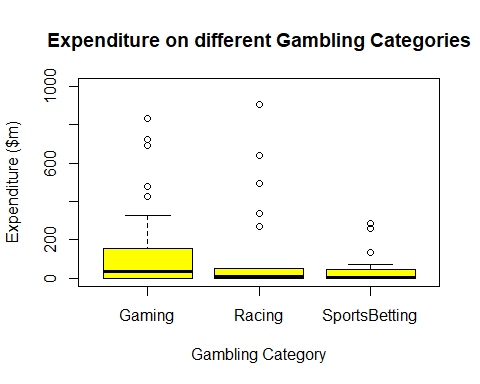
state.expendo.dist <- split(Gambling$Expenditure, Gambling$GamblingCategory)  
  
expp <- Gambling$Expenditure  
summary(expp)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.005 1.139 20.459 258.340 137.418 5744.291

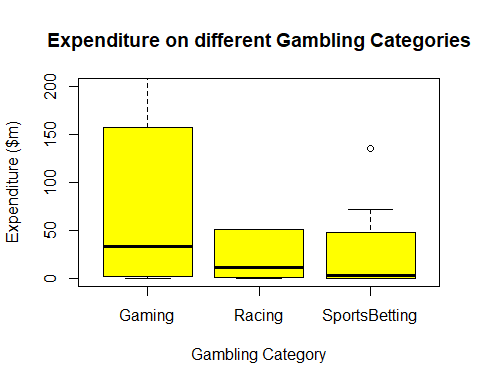
boxplot(state.expendo.dist, col = "yellow", ylim = c(0,6000), ylab = "Expenditure ($m)", xlab = "Gambling Category", main = "Expenditure on different Gambling Categories")



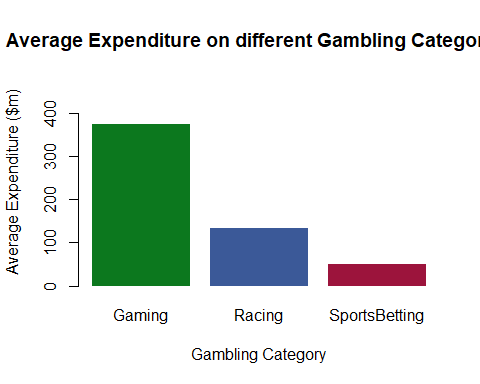
boxplot(state.expendo.dist, col = "yellow", ylim = c(0,1000), ylab = "Expenditure ($m)", xlab = "Gambling Category", main = "Expenditure on different Gambling Categories")



boxplot(state.expendo.dist, col = "yellow", ylim = c(0,200), ylab = "Expenditure ($m)", xlab = "Gambling Category", main = "Expenditure on different Gambling Categories")



Ns <- c(rgb(12,120,30, maxColorValue = 255), rgb(59,89,152, maxColorValue = 255), rgb(156,20,60, maxColorValue = 255))  
  
state.expendo.mean <- sapply(state.expendo.dist, mean)  
  
barplot(state.expendo.mean, ylim = c(0,480), col = Ns, border = NA, ylab = "Average Expenditure ($m)", xlab = "Gambling Category", main = "Average Expenditure on different Gambling Categories")



T = tail(state.expendo.mean, n=3)  
message ("The Expenditure values From Left to Right: ", toString(T))

## The Expenditure values From Left to Right: 374.600637254902, 134.034714285714, 50.912125

### Text and Analysis:

The highest value that was not an outliner in the gaming category was equivalent to $329.19m which was significantly higher than the maximum value for both the Racing $51.005m and the Sports Betting $71.451m categories. The minimum value for all three categories was approximately the same and was roughly equal to $ 0.001m. The median for the gaming category was higher than the medians of the Racing and Sports Betting categories. The bar plot represents the average expenditure on each one of the gambling categories, the highest average expenditure was for the gaming category and was equivalent to $374.6m while the lowest was on Sports Betting and was equivalent to $50.9m, also the racing category had an average expenditure of $134.9m.

### Summary:

According to the data set, there are three main gambling categories. Those categories are, Gaming, Racing and Sports betting. Gaming is the broadest gambling category and it’s subdivided into a large number of platforms. This includes but not limited to, Lotto, Keno, Pools, Casino etc. Racing consists of three main gambling types which are, TAB racing, On/Off course book maker and On/ Off course Totalisator. Finally, Sports betting and it also has three types, Book maker and other Fixed Odds, Tab Fixed Odds and Tab Tote Odds. This diversity can be clearly seen by looking at the boxplot in which the lengthy Gaming box indicates a large diversity of money expenditure while the equally-sized short boxes of Sports Betting and Racing indicates a smaller diversity. The median for all the gambling categories is relatively low which shows that 50% of all the money expenditure is situated within the range of 0 to approximately $40m. Additionally, the bar chart clearly shows that the average gambling expenditure on the Gaming category is the highest. In conclusion, it was found that gaming was the most diverse gambling category and was the one that produced the largest net income (expenditure). Also, racing came in the middle while sports betting came last.

# Related Article:

URL: <https://www.australianbetting.org/gambling-australias-obsession/>

The main idea that this article is revolving around is about the growing gambling addiction in Australia. The article states that the growing presence of different gambling platforms is a major reason behind the growing addiction. The article further explains that new laws and legislations have been implemented in different Australian states to curb the ever growing addiction. Finally, the article talks about the positive aspect of gambling as it provides jobs opportunities and generates large income that can help the countries’ economy. The article uses statistical data as an evidence that supports their claim, the data was plotted on a bar chart that shows the approximate adults lose each year to gambling in 10 different countries around the world. This article provides me with ideas of how to study the trends in the gambling data set in a more effective manner. That is, I was able to produce a number of questions around gambling that can be proved using the gambling data set as an evidence. For instance, which state is the largest contributor to the money spent on gambling and does the new laws and legislations make any impact on the growing gambling addiction?

# Conclusion:

In conclusion, the statistical data set provided a concrete evidence that enabled us to answer our two research questions accurately. Research question number one showed that the state of NSW had the highest gambling expenditure with the state of VIC been second. Additionally, it was observed that the increase in population within a state will significantly affect the overall gambling expenditure within that certain state. For the second research question, it was discovered that the gambling category “Gaming” was the most diverse category and had the highest average gambling expenditure while “Racing” was the second most diverse and had the second highest average gambling expenditure and finally “Sports Betting” was the last gambling category in both diversity and average gambling expenditure.