# **Data Analysis of Poverty**

**IBM Descriptive Analytics**

**(IBMDESC)**

**Professor: Mr. Joaquin Bamba**

**Belchez, Maica L.**

**Heramia, Johanna Marisse C.**

**Tadeo, Jose Lorenzo G.**

**March 28, 2017**

Table of Contents

[Data Analysis of Poverty 1](#_Toc478333050)

[Introduction 3](#_Toc478333051)

[Population Per Region 4](#_Toc478333052)

[Family Per Region 2012 7](#_Toc478333053)

[Family Per Region 2015 8](#_Toc478333054)

[Income of Families Per Region 2012 and 2015 10](#_Toc478333055)

[Expenditure of Families Per Region 12](#_Toc478333056)

[Poverty Incidence by Population 2012 13](#_Toc478333057)

[Poverty Incidence by Population 2015 14](#_Toc478333058)

[Poverty Incidence by Families 2012 and 2015 15](#_Toc478333059)

[Labor Productivity by Population 2012 and 2015 17](#_Toc478333060)

[Number of Schools by Population 19](#_Toc478333061)

[Annual Poverty Threshold by Population 20](#_Toc478333062)

[Recommendation and Conclusion 23](#_Toc478333063)

## **Introduction**

What is Poverty? Poverty, according to Randolf S. David, it is theInability to provide for the basic requirements of minimum survival – such as food, housing, clothing, medical care and education. The question we must ask is why large numbers of a country’s population find themselves in this situation. 26.3 percent of Filipinos were found to be living below the poverty line, per Philippine Statistics Authority. Poor people are getting poorer while rich are getting richer and there are many causes or factors why Filipinos are still under this never-ending poverty. They don’t earn enough money to buy their needs. This maybe because of the government doesn’t provide adequate social services to those who could least afford, there are many children in household, high inflation of prices because of the low economic growth, low job quality resulting in low income, lack of education, the recurring natural disasters, and more.

You will see statistical data of poverty in the Philippines year 2012 and 2015 in different regions through the graphical representation using R. Population, number of families, poverty incidence, family income and expenditure, Annual Per Capita Poverty Threshold per region and more important role that may help in analyzing the poverty and what are the factors contributing to it, how they are related to each other, its summaries and analysis.

## **Population Per Region**

The region that has highest population as of 2012 is in Region IV-A or CALABARZON, to be exact, it has 13,458,967. (Data below are rounded-off to hundreds for better view of data in graph.) This data will help us determine if percentage of poverty is related or has a positive or negative relationship to the number of population by region.

As of 2015, the region that has the highest population is not in Region IV-A or

CALABARZON anymore, but in CAR or Cordillera Administrative Region. As of the year 2015, the population for CAR increased by 15,545,389 while the previous highest population for region IV-A only increased by 955,807. Comparing it to another region’s increase in population, the increase in CAR’s population is the most noticeable.

> population2012=tapply(DataV3$Population2012, DataV3$RegionalDesignation, FUN=sum, na.rm=TRUE)

> population2012

ARMM CAR NCR Region I Region II Region III

33625.01 16766.17 123154.37 48744.00 33259.19 106096.16

Region IV-A Region IV-B Region IX Region V Region VI Region VII

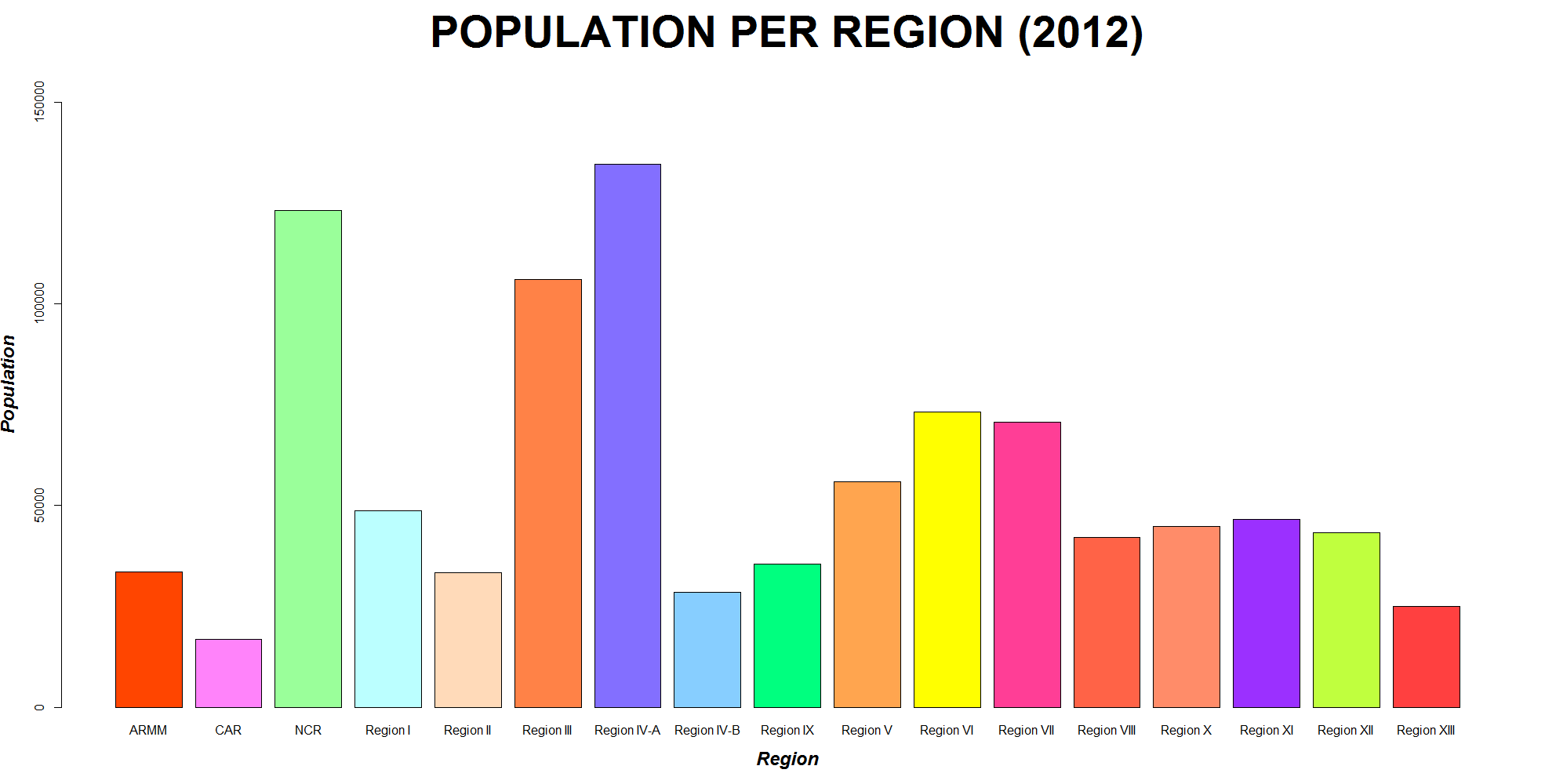
134589.67 28510.28 35456.10 55919.11 73091.53 70609.03

Region VIII Region X Region XI Region XII Region XIII

42146.33 44896.14 46600.27 43306.22 25074.10

> colors <- c("orangered1", "orchid1", "palegreen1", "paleturquoise1", "peachpuff1", "sienna1", "slateblue1", "skyblue1", "springgreen1", "tan1", "yellow1", "violetred1", "tomato1", "salmon1", "purple1", "olivedrab1", "brown1")

> barplot(population2012, ylab="Population", xlab="Region", ylim=c(0,160000),col=colors, main="POPULATION PER REGION (2012)", cex.lab=1.5, cex.main=3.5, font.lab=4, font.main=2)



> library(lattice)

> t4=tapply(DataV3$Population2015, DataV3$RegionalDesignation, mean, na.rm="TRUE")

> t4

ARMM CAR NCR Region I Region II Region III

37813.87 172220.06 128772.53 50261.28 34514.10 112181.77

Region IV-A Region IV-B Region IX Region V Region VI Region VII

144147.74 29633.60 36297.83 57969.89 44772.47 60419.03

Region VIII Region X Region XI Region XII Region XIII

44401.50 46893.02 48933.18 45452.76 25967.09

> t5=tapply(DataV3$PovertyIncidenceAmongPopulation2015, DataV3$RegionalDesignation, mean, na.rm="TRUE")

> t5

ARMM CAR NCR Region I Region II Region III

53.7 19.7 3.9 13.1 15.8 11.2

Region IV-A Region IV-B Region IX Region V Region VI Region VII

9.1 24.4 33.9 36 22.4 27.6

Region VIII Region X Region XI Region XII Region XIII

38.7 36.6 22 37.3 39.1

> t6 = cbind(t4,t5)

> t7=t6[t6[,1]>18,]

> t7[order(t7[,1], decreasing = TRUE),]

t4 t5

CAR 172220.06 19.7

Region IV-A 144147.74 9.1

NCR 128772.53 3.9

Region III 112181.77 11.2

Region VII 60419.03 27.6

Region V 57969.89 36.0

Region I 50261.28 13.1

Region XI 48933.18 22.0

Region X 46893.02 36.6

Region XII 45452.76 37.3

Region VI 44772.47 22.4

Region VIII 44401.50 38.7

ARMM 37813.87 53.7

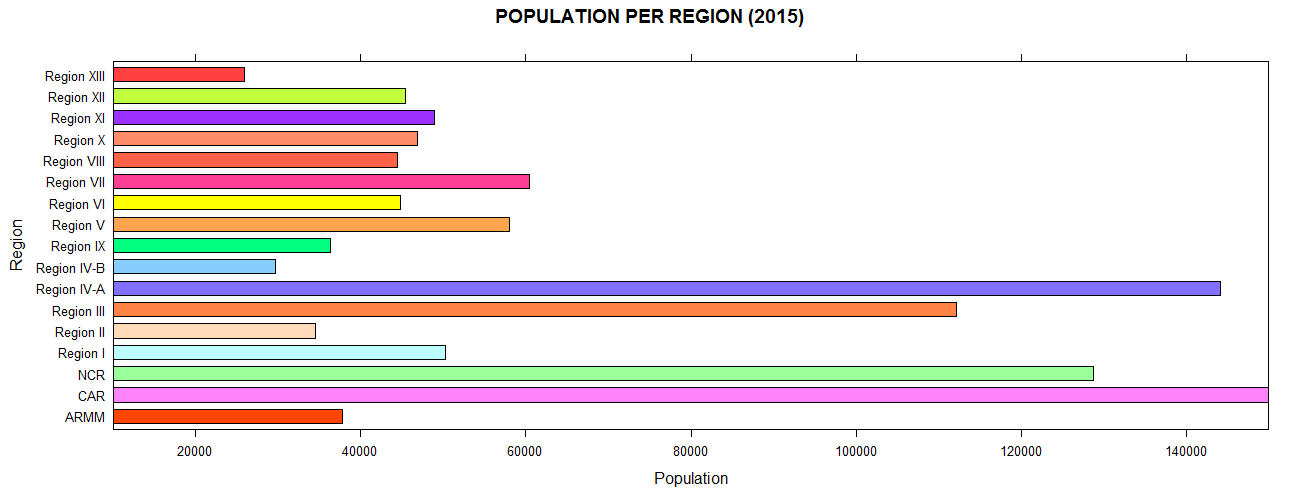
Region IX 36297.83 33.9

Region II 34514.10 15.8

Region IV-B 29633.60 24.4

Region XIII 25967.09 39.1

> colors <- c("orangered1", "orchid1", "palegreen1", "paleturquoise1", "peachpuff1", "sienna1", "slateblue1", "skyblue1", "springgreen1", "tan1", "yellow1", "violetred1", "tomato1", "salmon1", "purple1", "olivedrab1", "brown1")

> barchart(t7[,1], col=colors, xlab="Population", ylab="Region", main="POPULATION PER REGION (2015)", cex.lab=1.5, cex.main=3.5, font.lab=4, font.main=2, xlim=c(10000,150000))

## **Family Per Region 2012**

The region that has the highest number of families is Region IV-A or CALABARZON with a total of 3,082 families. The second highest total families in a region or close to the total number of families in Region IV-A is NCR or National Capital Region with a total number of 2,917 families. For the least number of families in a region, CAR has the fewest number of families with a total of 375 residing in that region comparing to another region’s number of families. After CAR, the next fewest number of families residing in a region is Region XIII with a total of 532.

> family2012=tapply(DataV3$NumberOfFamilies2012,DataV3$RegionalDesignation, FUN=sum, na.rm=TRUE)

> family2012

ARMM CAR NCR Region I Region II Region III

557 375 2917 1105 771 2386

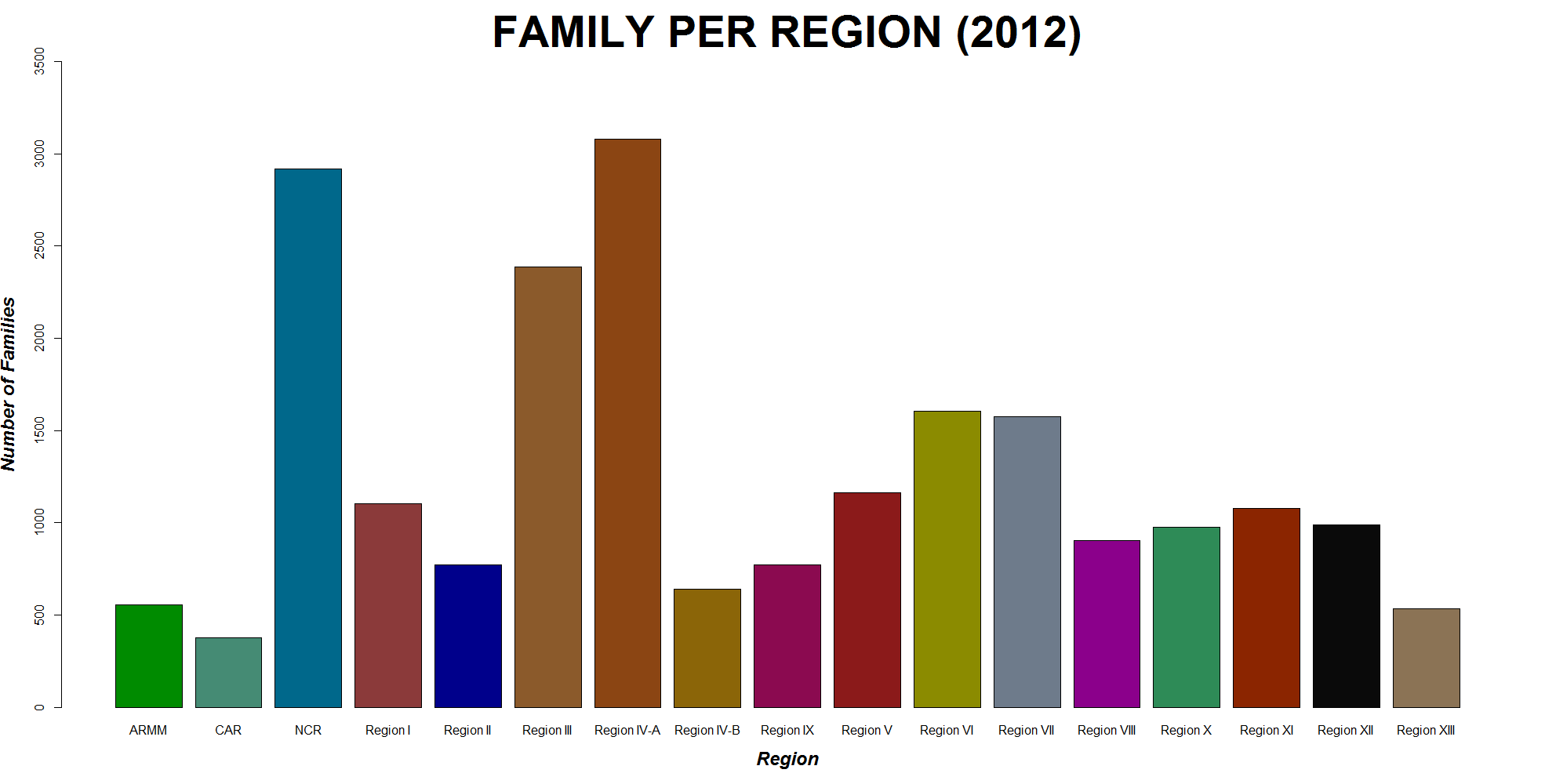
Region IV-A Region IV-B Region IX Region V Region VI Region VII

3082 638 772 1165 1604 1577

Region VIII Region X Region XI Region XII Region XIII

902 976 1078 988 532

> colors <- c("green4", "aquamarine4", "deepskyblue4", "indianred4", "blue4", "tan4", "chocolate4", "darkgoldenrod4", "deeppink4", "firebrick4", "yellow4", "lightsteelblue4", "magenta4", "seagreen4", "orangered4", "gray4", "burlywood4")

> barplot(family2012,ylab="Number of Families",xlab="Region",ylim=c(0,3500),col=colors,main="FAMILY PER REGION (2012)", cex.lab=1.5, cex.main=3.5, font.lab=4, font.main=2)

## **Family Per Region 2015**

As of 2015, the highest number of families in a region is still Region IV-A or CALABARZON and has increased in the past 3 years by 169 families. When you compare the data of “Family Per Region 2012” to “Family Per Region 2015”, you will see the increase in number of families residing in that region. The second highest number of families in a region is NCR with an increase of 102 in the past 3 years. The least number of families in a region is CAR with a total of 402.

> t4=tapply(DataV3$NumberOfFamilies2015, DataV3$RegionalDesignation, mean, na.rm="TRUE")

> t5=tapply(DataV3$PovertyIncidenceAmongFamilies2015, DataV3$RegionalDesignation, mean, na.rm="TRUE")

> t6 = cbind(t4,t5)

> t7=t6[t6[,1]>18,]

> t7[order(t7[,1], decreasing = TRUE),]

t4 t5

Region IV-A 3251 6.7

NCR 3019 2.7

Region III 2507 8.9

Region VI 1699 16.6

Region VII 1672 23.6

Region V 1262 27.5

Region I 1170 9.6

Region XI 1156 16.6

Region XII 1055 30.5

Region X 1029 30.3

Region VIII 976 30.7

Region IX 824 26.0

Region II 816 11.7

Region IV-B 697 17.4

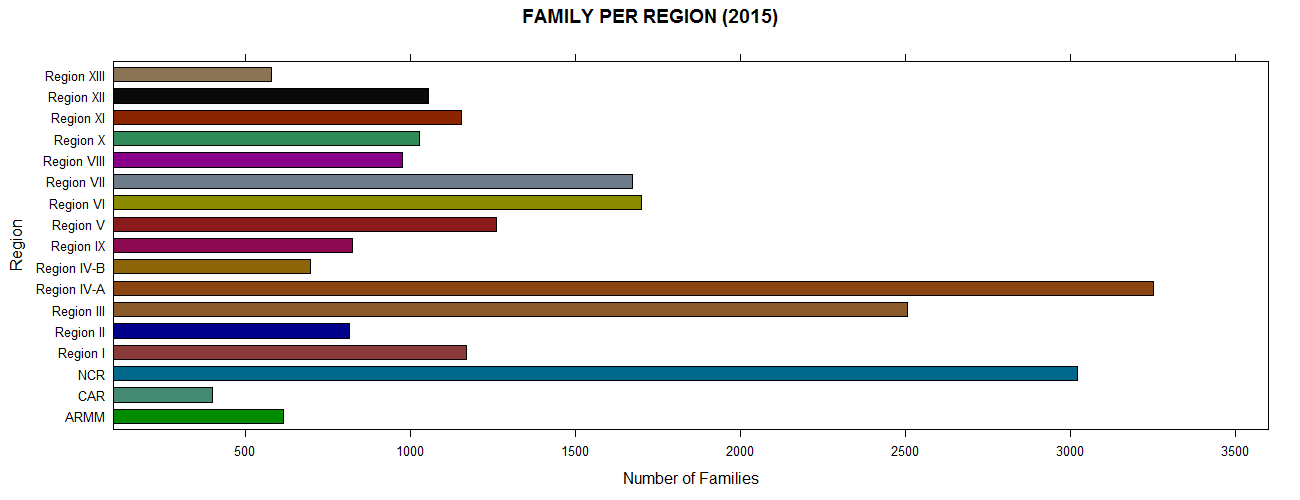
ARMM 616 48.2

Region XIII 579 30.8

CAR 402 14.4

> colors <- c("green4", "aquamarine4", "deepskyblue4", "indianred4", "blue4", "tan4", "chocolate4", "darkgoldenrod4", "deeppink4", "firebrick4", "yellow4", "lightsteelblue4", "magenta4", "seagreen4", "orangered4", "gray4", "burlywood4")

> barchart(t7[,1], col=colors, xlab="Number of Families", ylab="Region", main="FAMILY PER REGION (2015)", cex.lab=1.5, cex.main=3.5, font.lab=4, font.main=2, xlim=c(100,3600))



## **Income of Families Per Region 2012 and 2015**

In the year 2012, the highest income of families per region is NCR with a total of 1,106,169 and second to the highest is Region IV-A or CALABARZON with a total of 876,006. The fewest income of families per region is in ARMM with only a 72,196 and second to it is Region XIII with only a 95,809.

In the year 2015, the highest income of families per region is still NCR and second to it is still Region IV-A or CALABARZON. Comparing the increase in income for the past 3 years, NCR has the most increase with a total of 176,700. NCR is where all big companies reside. Giving opportunity to people that resides in NCR.

> Incomefamily2012=tapply(DataV3$TotalIncomeOfFamilies2012,DataV3$RegionalDesignation, FUN=sum, na.rm=TRUE)

> Incomefamily2012

ARMM CAR NCR Region I Region II Region III

721.96 963.51 11061.69 2250.50 1500.89 6188.93

Region IV-A Region IV-B Region IX Region V Region VI Region VII

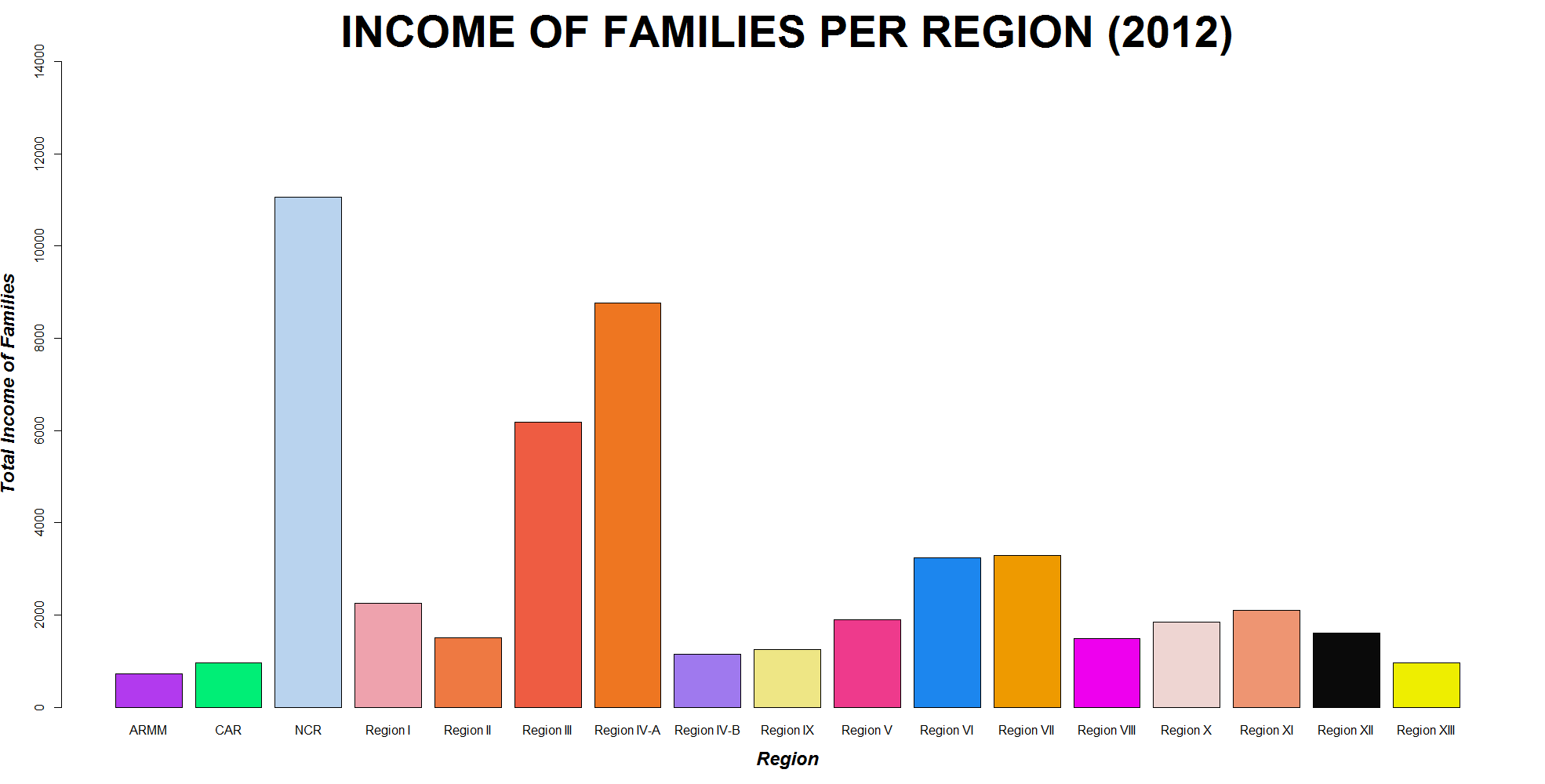
8760.06 1141.17 1249.03 1891.85 3240.28 3294.15

Region VIII Region X Region XI Region XII Region XIII

1494.93 1850.15 2094.05 1606.74 958.09

> colors <- c("darkorchid2", "springgreen2", "slategray2", "lightpink2", "sienna2", "tomato2", "chocolate2", "mediumpurple2", "khaki2", "violetred2", "dodgerblue2", "orange2", "magenta2", "mistyrose2", "lightsalmon2", "gray4", "yellow2")

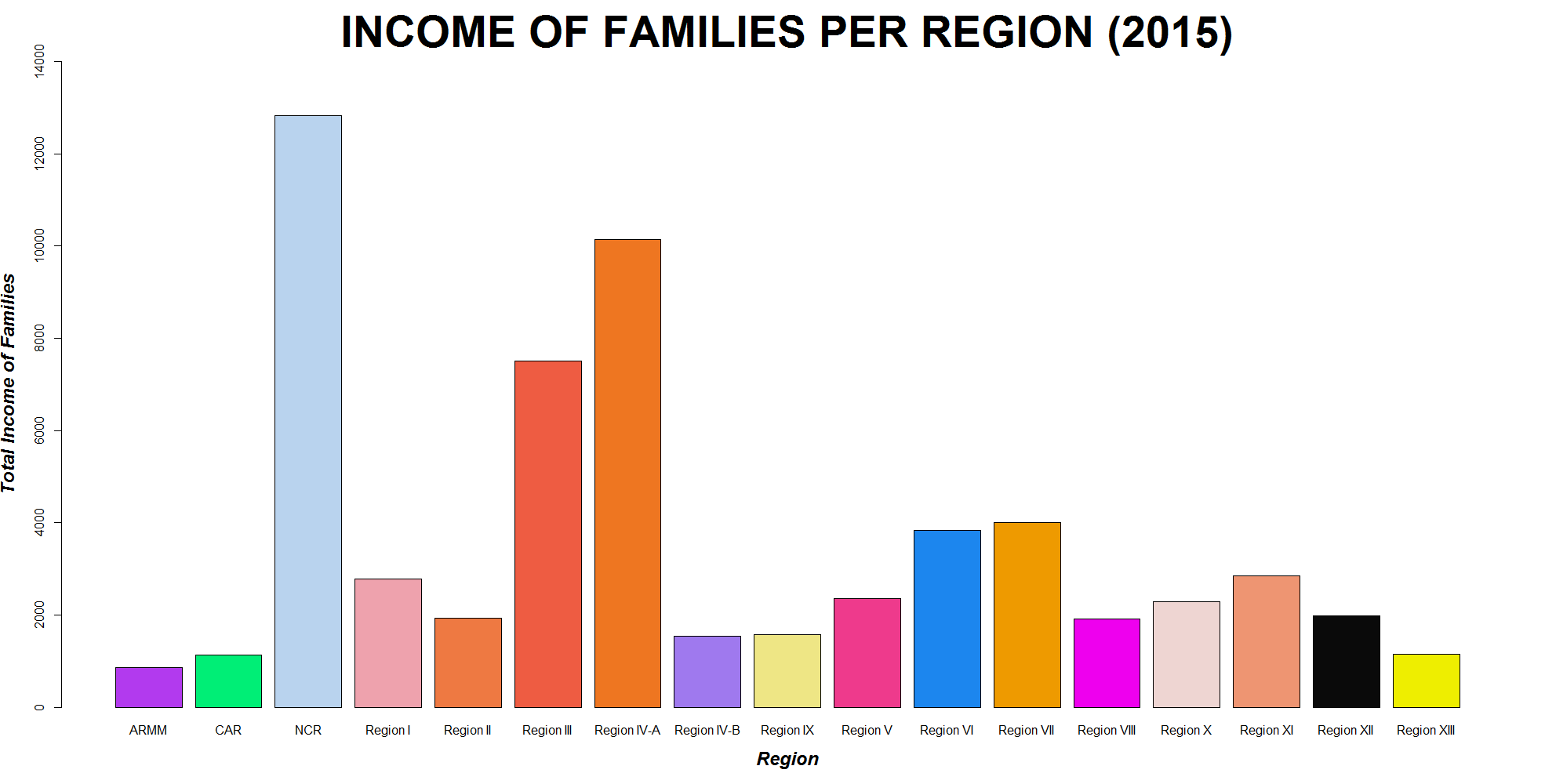
> barplot(Incomefamily2012,ylab="Total Income of Families",xlab="Region",col=colors,main="INCOME OF FAMILIES PER REGION (2012)", cex.lab=1.5, cex.main=3.5, font.lab=4, font.main=2, ylim=c(0,14000))



> Incomefamily2015=tapply(DataV3$TotalIncomeOfFamilies2015,DataV3$RegionalDesignation, FUN=sum, na.rm=TRUE)

> colors <- c("darkorchid2", "springgreen2", "slategray2", "lightpink2", "sienna2", "tomato2", "chocolate2", "mediumpurple2", "khaki2", "violetred2", "dodgerblue2", "orange2", "magenta2", "mistyrose2", "lightsalmon2", "gray4", "yellow2")

> barplot(Incomefamily2015,ylab="Total Income of Families",xlab="Region",col=colors,main="INCOME OF FAMILIES PER REGION (2015)", cex.lab=1.5, cex.main=3.5, font.lab=4, font.main=2, ylim=c(0,14000))



## **Expenditure of Families Per Region**

As of 2012, the highest expenditure of families per region is NCR with a total of 947,599. NCR has the most income therefore it also has the highest expenditure. Second to NCR is Region IV-A with a total of 748,129. Just like the graph in “Income of Families per Region” the higher the income, the higher expenditure of a family. Aside for NCR having the highest income per family. NCR is where all big shopping company is whether it is a local shopping company or international shopping company. That is why families in NCR has more expenditure than other regions.

> Expenditurefamily2012=tapply(DataV3$TotalExpenditureOfFamilies2012,DataV3$RegionalDesignation, FUN=sum, na.rm=TRUE)

> Expenditurefamily2012

ARMM CAR NCR Region I Region II Region III

635.06 706.33 9475.99 1752.20 1078.84 5039.95

Region IV-A Region IV-B Region IX Region V Region VI Region VII

7481.29 882.15 938.29 1672.56 2609.71 2586.35

Region VIII Region X Region XI Region XII Region XIII

1189.98 1393.53 1676.51 1380.89 753.52

> colors <- c("aliceblue", "cornflowerblue", "blueviolet", "darkcyan", "darkgray", "darkkhaki", "forestgreen", "greenyellow", "lightgreen", "midnightblue", "mediumvioletred", "moccasin", "maroon", "saddlebrown", "peru", "sandybrown", "yellowgreen")

> barplot(Expenditurefamily2012,ylab="Total Expenditure of Families",xlab="Region",col=colors,main="EXPENDITURE OF FAMILIES PER REGION (2012)", cex.lab=1.5, cex.main=3.5, font.lab=4, font.main=2, ylim=c(0,10000))



## **Poverty Incidence by Population 2012**

Poverty incidence is the number of families or population living below the poverty threshold. According to the representation, poverty incidence and population has negative relationship, which means that the higher the population, the lower the poverty incidence. How is that possible? We know that population is one of the factors that contributes to the poverty, and we must expect a positive relationship. Maybe there are other factors and later on we will see.

> plot(DataV3$Population2012, DataV3$PovertyIncidenceAmongPopulation2012, xlab="Population", ylab="Poverty Incidence", main="POVERTY INCIDENCE BY POPULATION (2012)", cex.lab=1.5, cex.main=2.5, font.lab=4, font.main=2, frame.plot=TRUE, pch=8, col="red4", xlim=c(20000,160000))

> poverpopu <- lm(DataV3$PovertyIncidenceAmongPopulation2012~DataV3$Population2012)

> abline(coef(poverpopu), lwd=2, col="sandybrown")



## **Poverty Incidence by Population 2015**

Poverty incidence is the number of families or population living below the poverty threshold. According to the representation, poverty incidence and population has negative relationship, which means that the higher the population, the lower the poverty incidence. How is that possible? We know that population is one of the factors that contributes to the poverty, and we must expect a positive relationship. Maybe there are other factors and later on we will see.

> plot(DataV3$Population2015, DataV3$PovertyIncidenceAmongPopulation2015, xlab="Population", ylab="Poverty Incidence", main="POVERTY INCIDENCE BY POPULATION (2015)", cex.lab=1.5, cex.main=2.5, font.lab=4, font.main=2, frame.plot=TRUE, pch=8, col="red4", xlim=c(20000,160000))

> poverpopu2 <- lm(DataV3$PovertyIncidenceAmongPopulation2015~DataV3$Population2015)

> abline(coef(poverpopu2), lwd=2, col="sandybrown")



## **Difference between Poverty Incidence by Population by Year**

The difference between the year 2012 and 2015 in poverty incidence by population per year is that in 2012, the highest poverty incidence is in ARMM exactly at 53.7% and still in 2015 the poverty incidence in that region did not decrease. The lowest poverty incidence is in NCR exactly at 3.9% and in 2015 like ARMM, the poverty incidence in that region did not decrease. The region that has the most decrease in poverty incidence is Region XI. In the year 2012, Region XI’s poverty incidence has a total of 30.7%, however after 3 years, Region XI’s poverty incidence decreased by 8.7% and in 2015 Region XI’s poverty incidence is 22%.

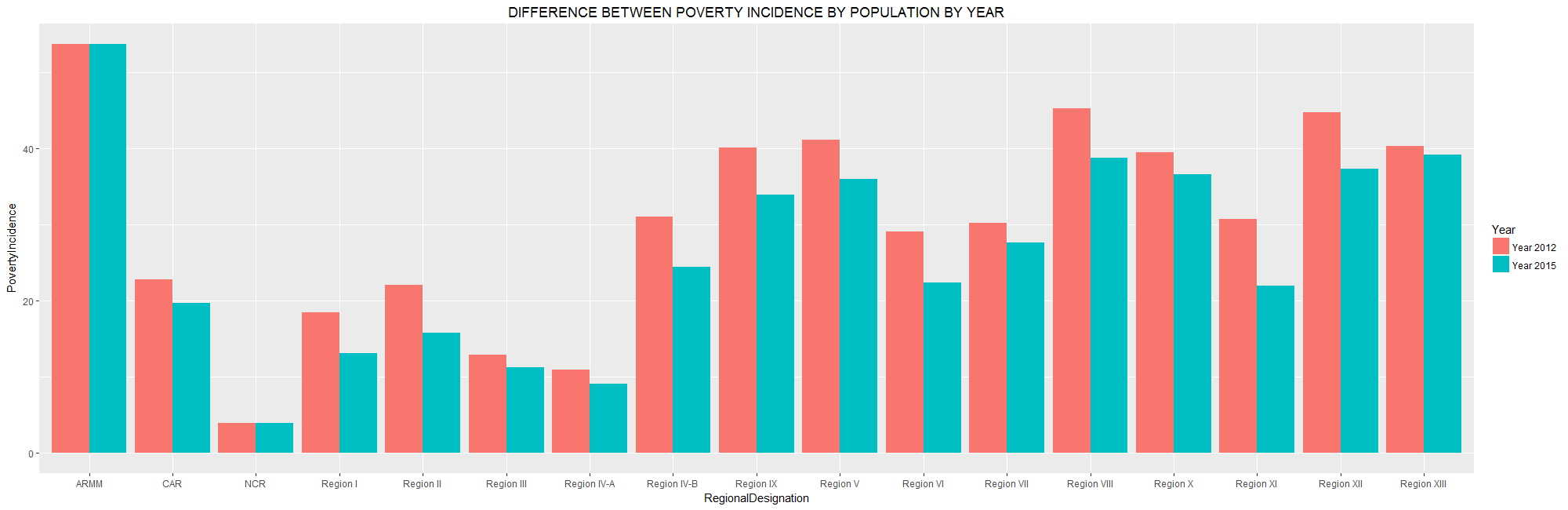
> library(readxl)

> ForGroupedBars <- read\_excel("C:/Users/johan/Desktop/IBMDESC-Poverty/Documentation/ForGroupedBars.xlsx")

> View(ForGroupedBars)

> library(ggplot2)

> ggplot(ForGroupedBars, aes(x=RegionalDesignation, y=PovertyIncidence,fill=Year)) + geom\_bar(stat = "identity",position=position\_dodge()) + ggtitle("DIFFERENCE BETWEEN POVERTY INCIDENCE BY POPULATION BY YEAR") + theme(plot.title = element\_text(hjust = 0.5))



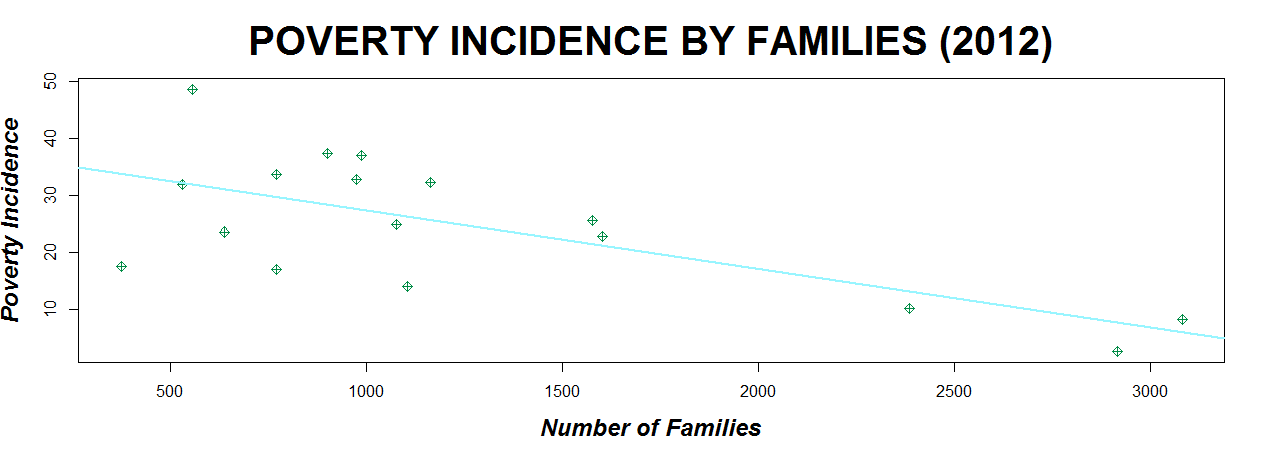
## **Poverty Incidence by Families 2012 and 2015**

The same here in number of families, the higher the number of families, the lower the poverty incidence both in 2012 and 2015

> plot(DataV3$NumberOfFamilies2012, DataV3$PovertyIncidenceAmongFamilies2012, xlab="Number of Families", ylab="Poverty Incidence", main="POVERTY INCIDENCE BY FAMILIES (2012)", cex.lab=1.5, cex.main=2.5, font.lab=4, font.main=2, frame.plot=TRUE, pch=9, col="springgreen4")

> poverfami2012 <- lm(DataV3$PovertyIncidenceAmongFamilies2012~DataV3$NumberOfFamilies2012)

> abline(coef(poverfami2012), lwd=2, col="cadetblue1")



> plot(DataV3$NumberOfFamilies2015, DataV3$PovertyIncidenceAmongFamilies2015, xlab="Number of Families", ylab="Poverty Incidence", main="POVERTY INCIDENCE BY FAMILIES (2015)", cex.lab=1.5, cex.main=2.5, font.lab=4, font.main=2, frame.plot=TRUE, pch=9, col="springgreen4")

> poverfami2 <- lm(DataV3$PovertyIncidenceAmongFamilies2015~DataV3$NumberOfFamilies2015)

> abline(coef(poverfami2), lwd=2, col="cadetblue1")



## **Labor Productivity by Population 2012 and 2015**

Labor Productivity is a measure of economic growth within a country. Labor productivity measures the amount of goods and services produced by one hour of labor.

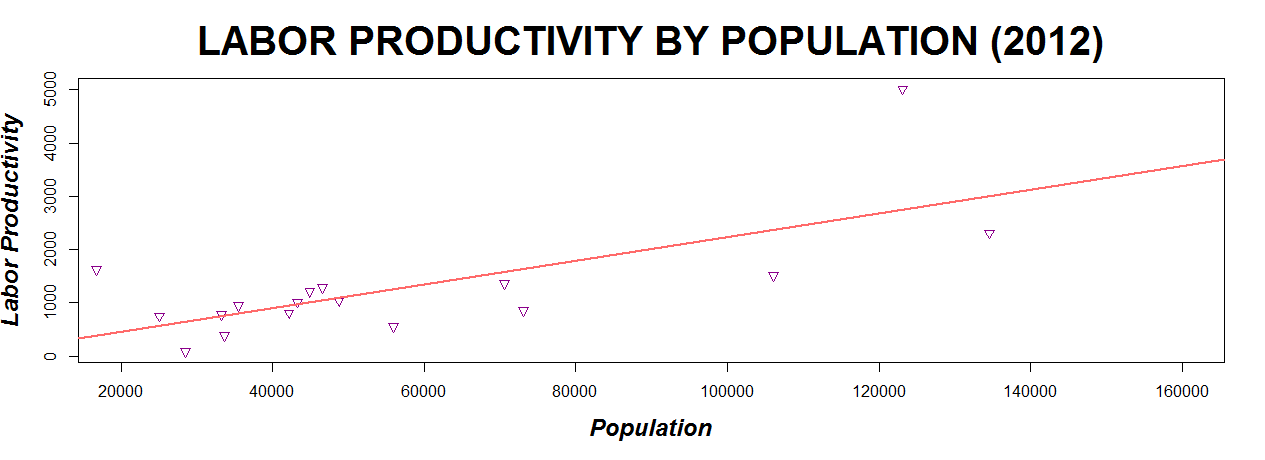
On the previous graphs, it is shown that the higher the population the lower the poverty incidence? Why is that? Here, in the graph shown below, we can see that there is a positive relationship between the labor productivity and population, which means that in every region of the Philippines, the higher the population, the higher the labor productivity is.

Now, in relation to poverty, as long as each person in the large population or large number of family is highly productive, it will help the economy or lessen the percentage of poverty. The reason why there is a negative relationship between the population and poverty incidence (as shown in the previous chart), is because of the high level of productivity of the population.

> plot(DataV3$Population2012, DataV3$LaborProductivity2012, xlab="Population", ylab="Labor Productivity", main="LABOR PRODUCTIVITY BY POPULATION (2012)", cex.lab=1.5, cex.main=2.5, font.lab=4, font.main=2, frame.plot=TRUE, pch=6, col="darkmagenta", xlim=c(20000,160000))

> laborpro2012 <-lm(DataV3$LaborProductivity2012~DataV3$Population2012)

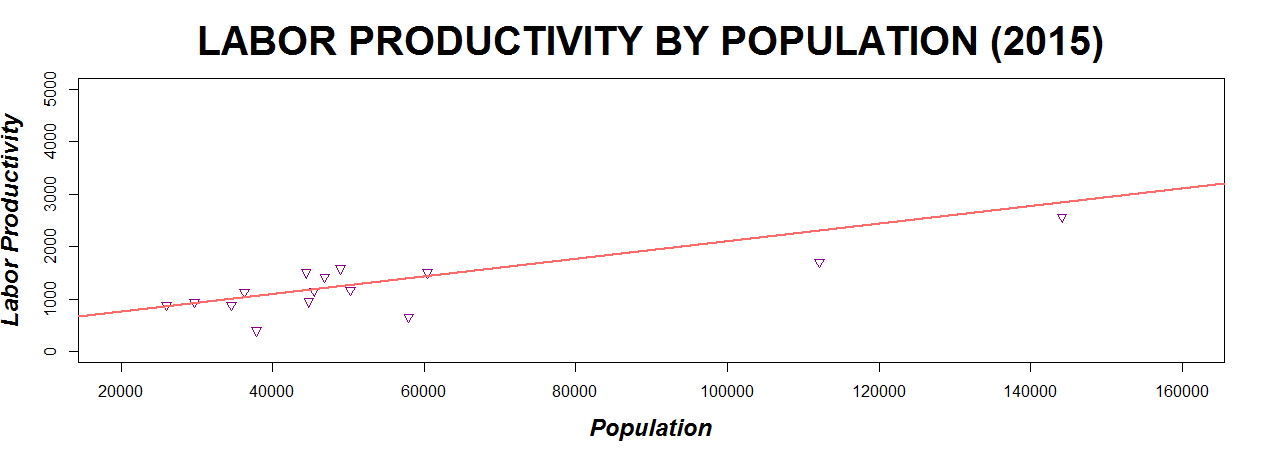
> abline(coef(laborpro2012), lwd=2, col="indianred1")



> plot(DataV3$Population2015, DataV3$LaborProductivity2015, xlab="Population", ylab="Labor Productivity", main="LABOR PRODUCTIVITY BY POPULATION (2015)", cex.lab=1.5, cex.main=2.5, font.lab=4, font.main=2, frame.plot=TRUE, pch=6, col="darkmagenta", ylim=c(0,5000), xlim=c(20000,160000))

> laborpro2015 <-lm(DataV3$LaborProductivity2015~DataV3$Population2015)

> abline(coef(laborpro2015), lwd=2, col="indianred1")



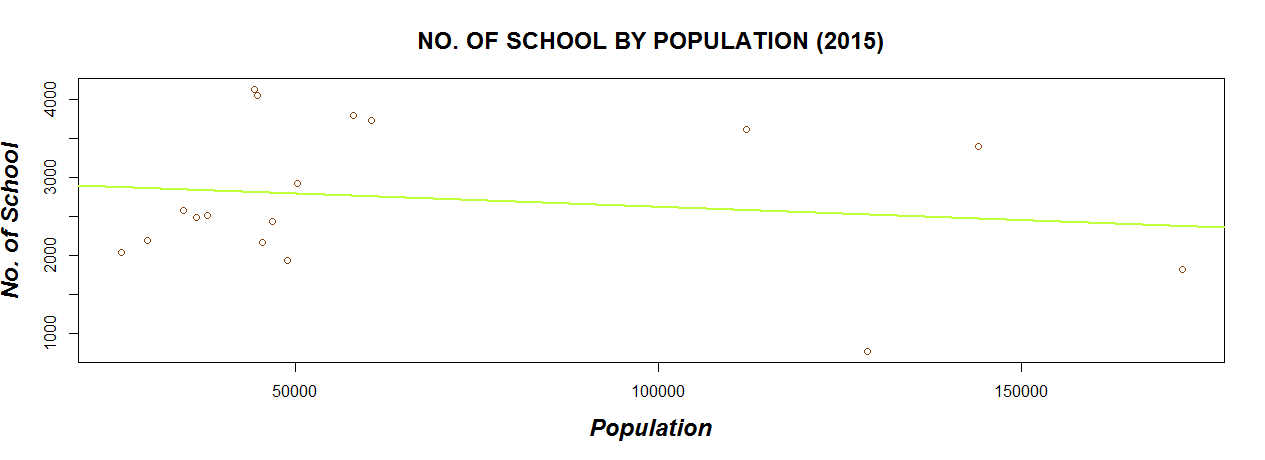
## **Number of Schools by Population**

As of 2015, the higher the population is the lower the number of school, but as you can see in the trendline, there is no to negative relationship between the number of school and population. This is also one of the factors that affects poverty and government must build more schools, not just to increase the employment rate but also to give education to more larger population.

> plot(DataV3$Population2015, DataV3$NoOfSchool, xlab="Population", ylab="No. of School", main="NO. OF SCHOOL BY POPULATION (2015)", cex.lab=1.5, cex.main=1.5, font.lab=4, font.main=2, frame.plot=TRUE, pch=1, col="saddlebrown")

> oOfSchool <- lm(DataV3$NoOfSchool~DataV3$Population2015)

> abline(coef(oOfSchool), lwd=2, col="olivedrab1")



## **Annual Poverty Threshold by Population**

Poverty thresholds are the original version of the measurement of poverty or it is the income required or the amount to be spent to satisfy the basic needs. How are we going to know if a person is below the poverty threshold or below the poverty line?

Example (from *Philippine Health Insurance Corporation*):

**Given:**

Family A, from Pasay City has seven (7) members, three (3) members are working with the following annual income:

|  |  |  |
| --- | --- | --- |
| Family Members | Occupation | Annual Income |
| Father | Farmer | ₱ 15, 000.00 |
| Mother | Housewife | - |
| Daughter A | Maid | * 18, 000.00 |
| Son A | Laborer | ₱ 36, 000.00 |
| Son B | Unemployed | - |
| Daughter B | Student | - |
| Grandmother | Unemployed | - |
| TOTAL ANNUAL FAMILY INCOME | | ₱ 69, 000.00 |

**Computation:**

**Total Annual Family Income ₱ 69, 000.00  
 No. of Family Members 7**

**₱ 9, 857.14**

**=**

**=**

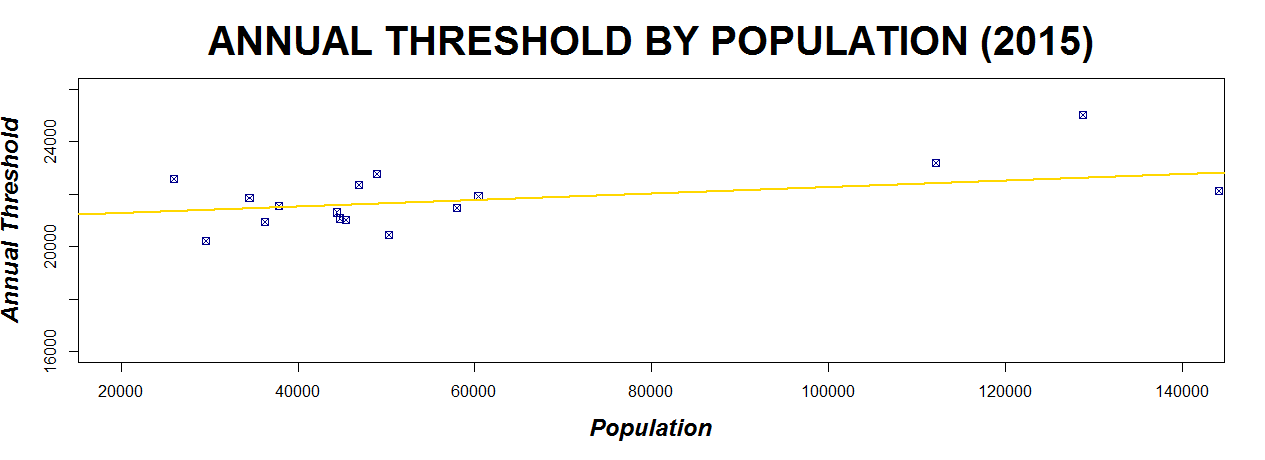
Pasay City is under the National Capital Region which has ₱ 23, 344.00 (as of 2012). If the **Annual Per Capita Income** is less than the **Annual Per Capita Poverty Threshold** of the region where the household is located, then the family is considered poor or indigent.

In the above example, the Annual Per Capita Income which is ₱ 9, 857.14 of Family A is evidently lower than Threshold Income of the region (₱23, 344.00). Therefore, Family A is considered indigent or poor.

> plot(DataV3$Population2015, DataV3$AnnualPerCapitaPovertyThreshold2015, xlab="Population", ylab="Annual Threshold", main="ANNUAL THRESHOLD BY POPULATION (2015)", cex.lab=1.5, cex.main=2.5, font.lab=4, font.main=2, frame.plot=TRUE, pch=7, col="darkblue", xlim=c(20000,140000), ylim=c(16000,26000))

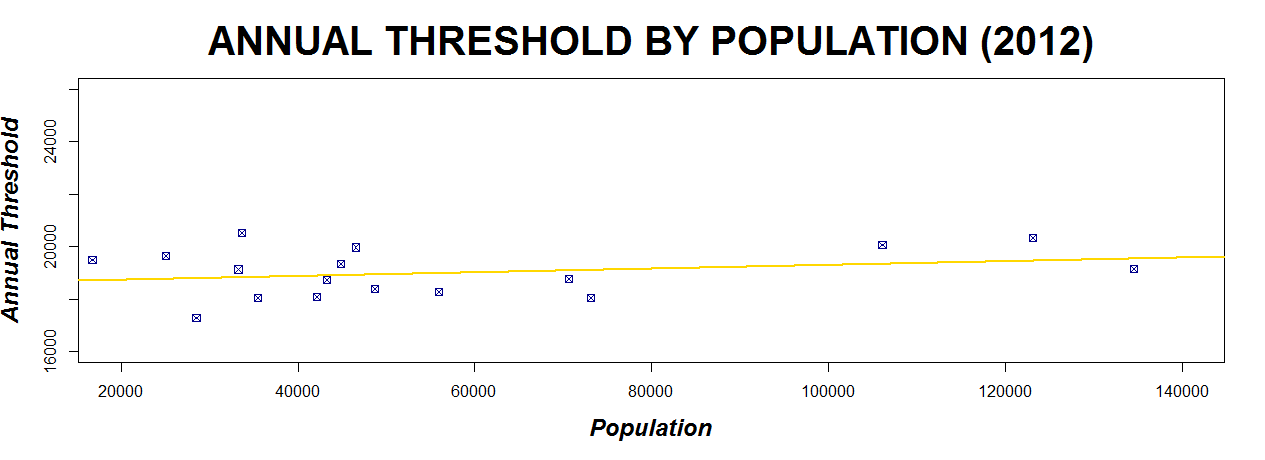
> annualthresh <- lm(DataV3$AnnualPerCapitaPovertyThreshold2015~DataV3$Population2015)

> abline(coef(annualthresh), lwd=2, col="gold")



> plot(DataV3$Population2012, DataV3$AnnualPerCapitaPovertyThreshold2012, xlab="Population", ylab="Annual Threshold", main="ANNUAL THRESHOLD BY POPULATION (2012)", cex.lab=1.5, cex.main=2.5, font.lab=4, font.main=2, frame.plot=TRUE, pch=7, col="darkblue", xlim=c(20000,140000), ylim=c(16000,26000))

> annualthresh <- lm(DataV3$AnnualPerCapitaPovertyThreshold2012~DataV3$Population2012)

> abline(coef(annualthresh), lwd=2, col="gold")

## **Recommendation and Conclusion**

Throughout the Philippines, poverty incidence among Filipinos registered at 26.3% as of 2015. Study suggests that building more schools, investing, building more infrastructure in the Philippines for more employment can help to alleviate the poverty. Based on the graphs or representation that was shown, population, education, productivity, etc., are the common factors that affect the poverty.

Population, factor that affects poverty, must be controlled by reinforcing more the family planning to households. Even if on the previous data shows that the higher population has lower poverty incidence because of the labor productivity, we must still control the population. What if the economy of the Philippines get low? The unemployment rate will increase, therefore increasing the number of productivity which is not good.

In addition, it is not enough that the government or other people that only helps to lessen the poverty. We, ourselves, must change. Change the way we think, the way we live, the way we accept change and the way we interact with people. Change should start from within, then change will happen around you.