# **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](#_Toc478138402)

[Introduction 3](#_Toc478138403)

[Population Per Region 4](#_Toc478138404)

[Poverty Incidence among Population Per Region 6](#_Toc478138405)

[Poverty Incidence among Family Per Region 8](#_Toc478138406)

[Poverty Incidence Per Region for 2012 and 2015 10](#_Toc478138407)

[Family Per Region 11](#_Toc478138408)

[Income of Families Per Region 12](#_Toc478138409)

[Expenditure of Families Per Region 13](#_Toc478138410)

[Annual Per Capita Poverty Threshold 14](#_Toc478138411)

[Poverty Incidence by Population 15](#_Toc478138412)

[Poverty Incidence by Population 16](#_Toc478138413)

[Poverty Incidence by Families 17](#_Toc478138414)

[Poverty Incidence by Families 18](#_Toc478138415)

[Recommendation and Conclusion 19](#_Toc478138416)

## **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, and medical care. 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 and the income are still low, lack of education and more.

You will see statistical data of poverty in the Philippines in different region 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 to the number of population by region.

> 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(0,200000),col=colors, main="POPULATION PER REGION (2012)", cex.lab=1.5, cex.main=3.5, font.lab=4, font.main=2)



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

CALABARZON, 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.

> population2015=tapply(DataV3$Population2015, DataV3$RegionalDesignation, FUN=sum, na.rm=TRUE)

> population2015

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

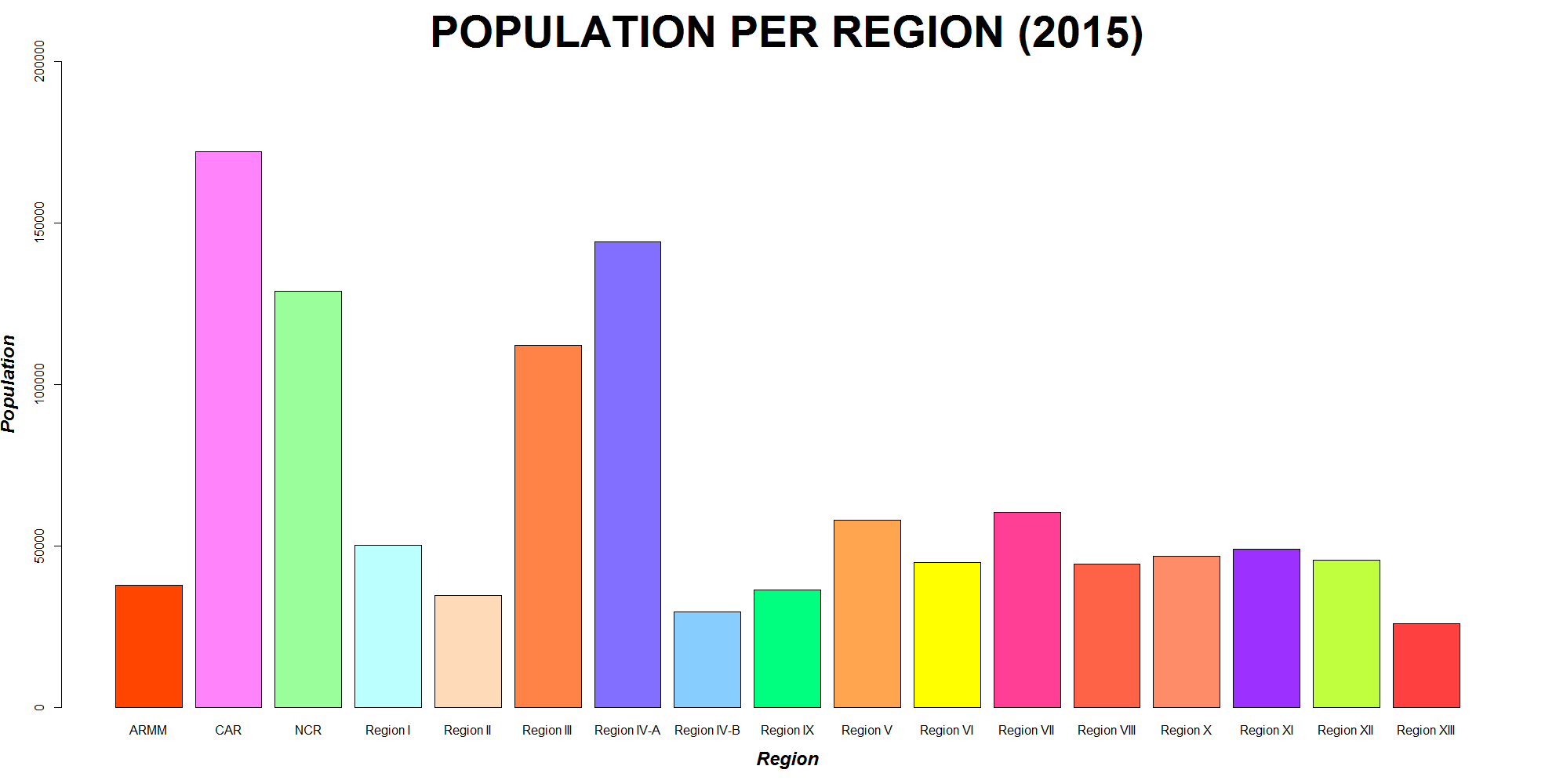
144147.74 29633.60 36297.83 57969.89 44772.47 60419.03

Region VIII Region X Region XI Region XII Region XIII

44401.50 46893.02 48933.18 45452.76 25967.09

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

> barplot(population2015, ylab="Population", xlab="Region", ylim=c(0,200000),col=colors, main="POPULATION PER REGION (2015)", cex.lab=1.5, cex.main=3.5, font.lab=4, font.main=2)



## **Poverty Incidence among Population Per Region 2012**

> t4=tapply(DataV3$Population2012, DataV3$RegionalDesignation, mean, na.rm="TRUE")

> t4

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

> t5=tapply(DataV3$PovertyIncidenceAmongPopulation2012, DataV3$RegionalDesignation, mean, na.rm="TRUE")

> t5

ARMM CAR NCR Region I Region II Region III

53.7 22.8 3.9 18.5 22.1 12.9

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

10.9 31.0 40.1 41.1 29.1 30.2

Region VIII Region X Region XI Region XII Region XIII

45.2 39.5 30.7 44.7 40.3

> t6 = cbind(t4,t5)

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

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

t4 t5

Region IV-A 134589.67 10.9

NCR 123154.37 3.9

Region III 106096.16 12.9

Region VI 73091.53 29.1

Region VII 70609.03 30.2

Region V 55919.11 41.1

Region I 48744.00 18.5

Region XI 46600.27 30.7

Region X 44896.14 39.5

Region XII 43306.22 44.7

Region VIII 42146.33 45.2

Region IX 35456.10 40.1

ARMM 33625.01 53.7

Region II 33259.19 22.1

Region IV-B 28510.28 31.0

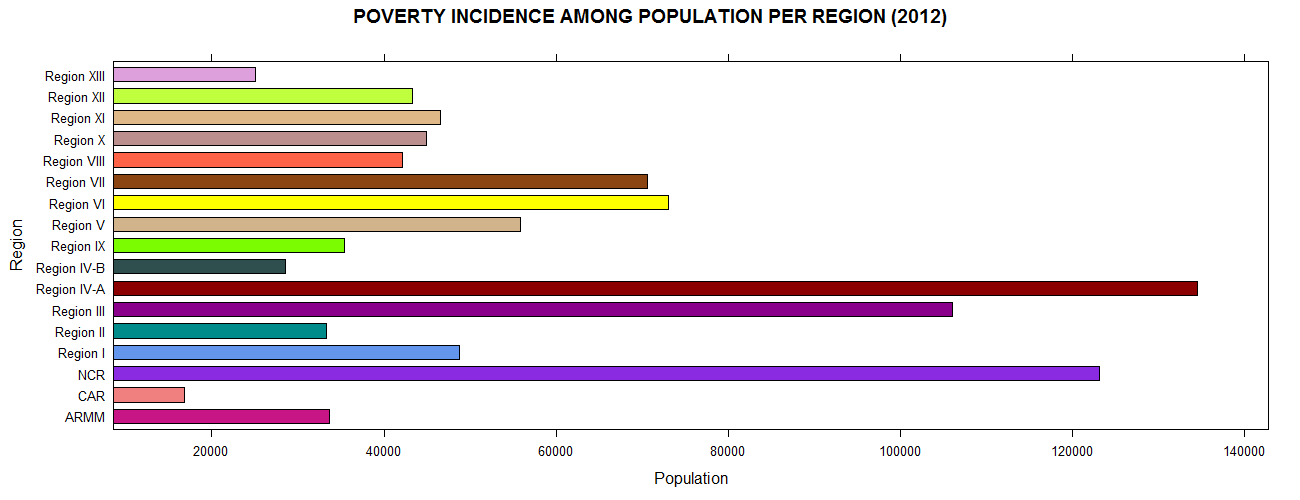
Region XIII 25074.10 40.3

CAR 16766.17 22.8

> library(lattice)

> colors <- c("mediumvioletred", "lightcoral", "blueviolet", "cornflowerblue", "darkcyan", "darkmagenta", "darkred", "darkslategray", "lawngreen", "tan", "yellow", "saddlebrown", "tomato", "rosybrown", "burlywood", "olivedrab1", "plum")

> barchart(t7[,1], col=colors, xlab="Population", ylab="Region", main="POVERTY INCIDENCE AMONG POPULATION PER REGION (2012)", cex.lab=1.5, cex.main=3.5, font.lab=4, font.main=2)



## **Poverty Incidence among Population 2015**

> 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("mediumvioletred", "lightcoral", "blueviolet", "cornflowerblue", "darkcyan", "darkmagenta", "darkred", "darkslategray", "lawngreen", "tan", "yellow", "saddlebrown", "tomato", "rosybrown", "burlywood", "olivedrab1", "plum")

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

## C:\Users\mlbel\AppData\Local\Microsoft\Windows\INetCacheContent.Word\PovIncPopPerReg2015.png

## **Poverty Incidence among Family Per Region 2012**

> library(lattice)

> t4=tapply(DataV3$NumberOfFamilies2012, DataV3$RegionalDesignation, mean, na.rm="TRUE")

> t4

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

> t5=tapply(DataV3$PovertyIncidenceAmongFamilies2012, DataV3$RegionalDesignation, mean, na.rm="TRUE")

> t5

ARMM CAR NCR Region I Region II Region III

48.7 17.5 2.6 14.0 17.0 10.1

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

8.3 23.6 33.7 32.3 22.8 25.7

Region VIII Region X Region XI Region XII Region XIII

37.4 32.8 25.0 37.1 31.9

> t6 = cbind(t4,t5)

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

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

t4 t5

Region IV-A 3082 8.3

NCR 2917 2.6

Region III 2386 10.1

Region VI 1604 22.8

Region VII 1577 25.7

Region V 1165 32.3

Region I 1105 14.0

Region XI 1078 25.0

Region XII 988 37.1

Region X 976 32.8

Region VIII 902 37.4

Region IX 772 33.7

Region II 771 17.0

Region IV-B 638 23.6

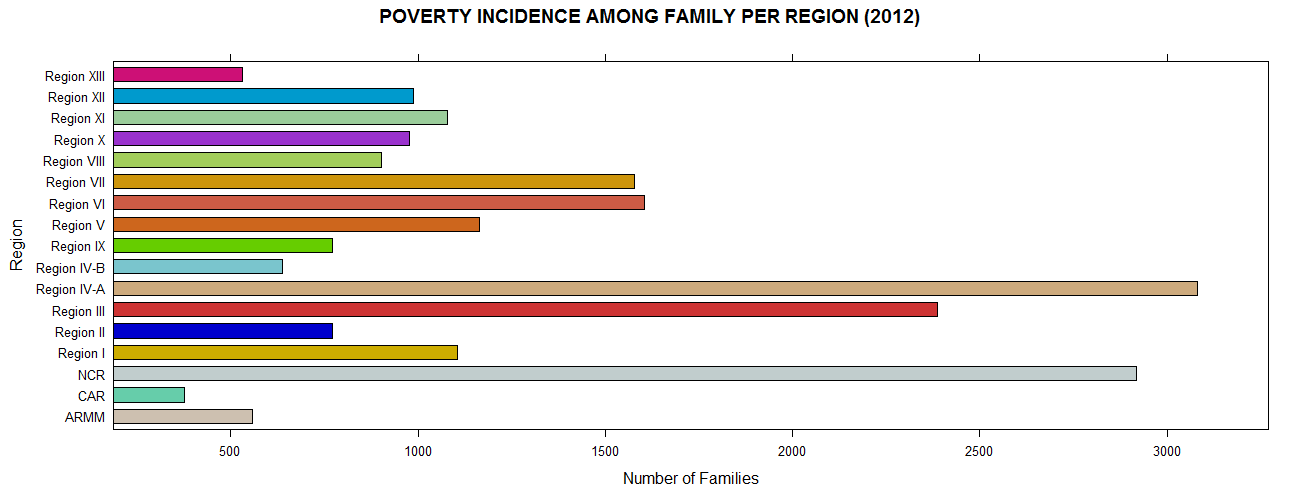
ARMM 557 48.7

Region XIII 532 31.9

CAR 375 17.5

> colors <- c("antiquewhite3", "aquamarine3", "azure3", "gold3", "blue3", "brown3", "burlywood3", "cadetblue3", "chartreuse3", "chocolate3", "coral3", "darkgoldenrod3", "darkolivegreen3", "darkorchid3", "darkseagreen3", "deepskyblue3", "deeppink3")

> barchart(t7[,1], col=colors, xlab="Number of Families", ylab="Region", main="POVERTY INCIDENCE AMONG FAMILY PER REGION (2012)", cex.lab=1.5, cex.main=3.5, font.lab=4, font.main=2)



## **Poverty Incidence among Family Per Region 2015**

> 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

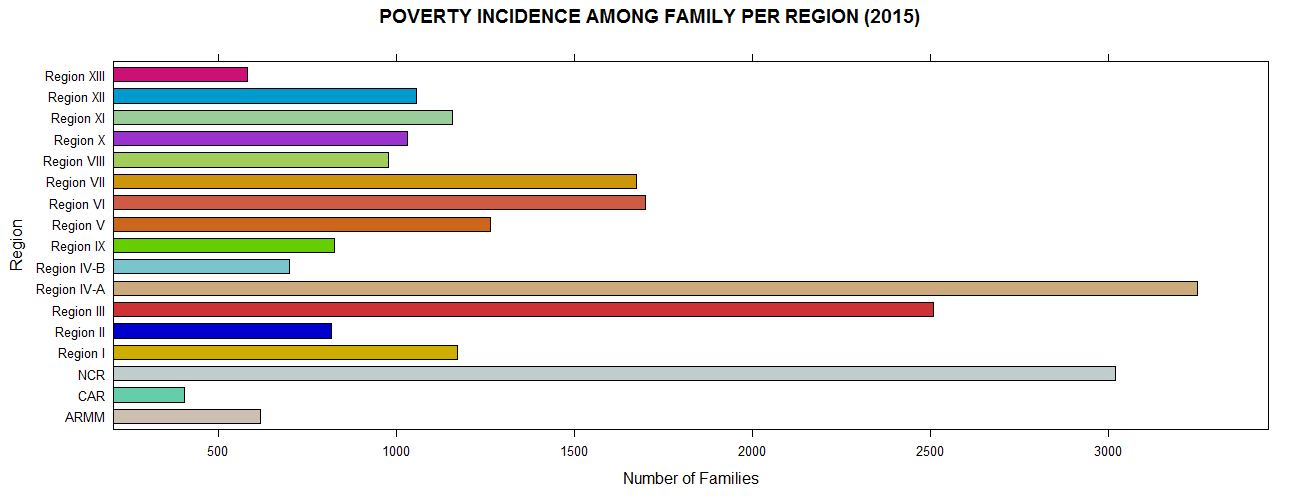
ARMM 616 48.2

Region XIII 579 30.8

CAR 402 14.4

> colors <- c("antiquewhite3", "aquamarine3", "azure3", "gold3", "blue3", "brown3", "burlywood3", "cadetblue3", "chartreuse3", "chocolate3", "coral3", "darkgoldenrod3", "darkolivegreen3", "darkorchid3", "darkseagreen3", "deepskyblue3", "deeppink3")

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



## **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)

## C:\Users\mlbel\AppData\Local\Microsoft\Windows\INetCacheContent.Word\FamPerReg2012 v2.png**Income of Families Per Region 2012**

> 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

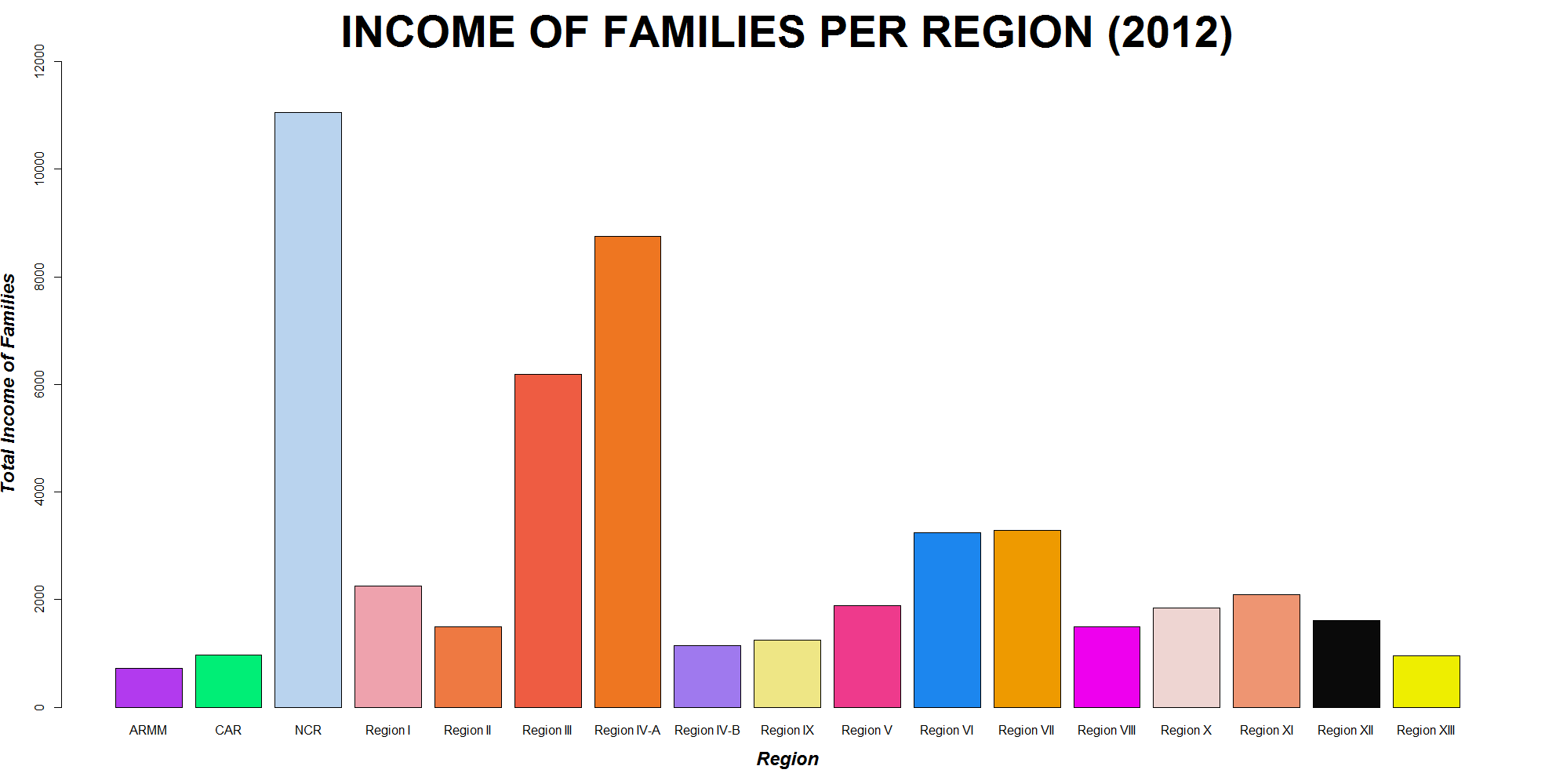
8760.06 1141.17 1249.03 1891.85 3240.28 3294.15

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(0,12000))



## **Expenditure of Families Per Region**

> 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**

> plot(Data$Population2015, Data$PovertyIncidenceAmongPopulation2015, xlab="Population", ylab="Poverty Incidence", main="Poverty Incidence By Population (2015)")

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

> abline(coef(poverpopu2), lwd=2)



## **Poverty Incidence by Population**

> plot(Data$Population2012, Data$PovertyIncidenceAmongPopulation2012, xlab="Population", ylab="Poverty Incidence", main="Poverty Incidence By Population (2012)")

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

> abline(coef(poverpopu), lwd=2)

## **Poverty Incidence by Families**

> plot(Data$NumberOfFamilies2015, Data$PovertyIncidenceAmongFamilies2015, xlab="Number Of Families", ylab="Poverty Incidence", main="Poverty Incidence By Families (2015)")

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

> abline(coef(poverfami2), lwd=2)

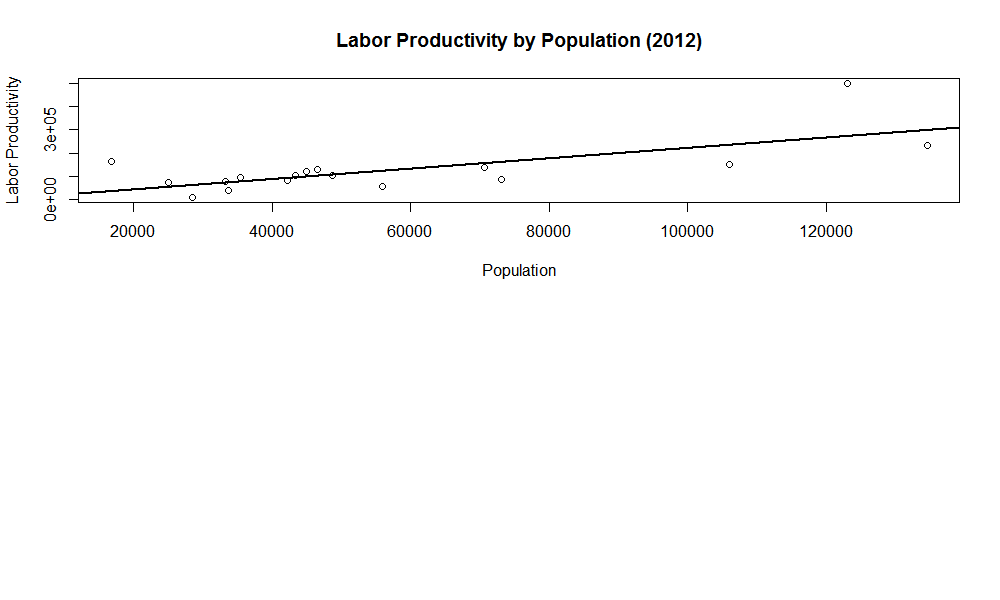


## **Labor Productivity by Population**

> plot(DataV3$Population2012, DataV3$LaborProductivity2012, xlab = "Population", ylab = "Labor Productivity", main = "Labor Productivity by Population (2012)")

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

> abline(coef(laborpro2012), lwd=2)

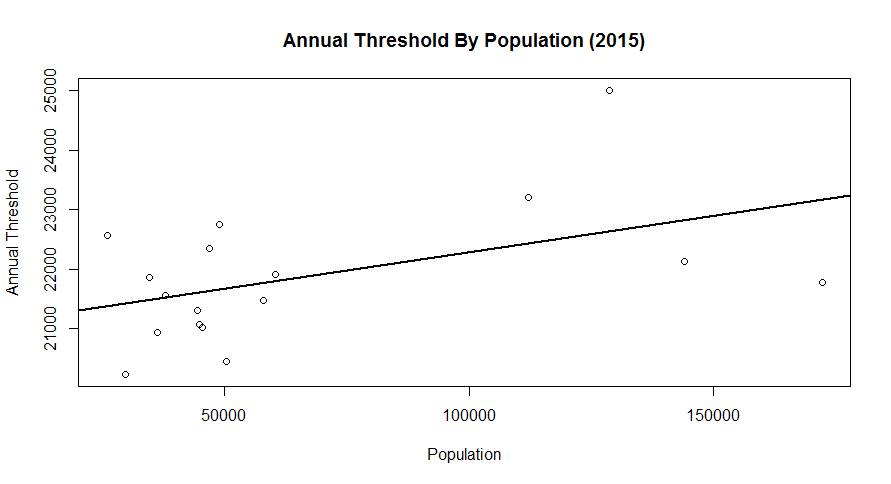


## **Annual Threshold by Population**

> plot(DataV3$Population2015, DataV3$AnnualPerCapitaPovertyThreshold2015, xlab="Population", ylab="Annual Threshold", main="Annual Threshold By Population (2015)")

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

> abline(coef(annualthresh), lwd=2)



## **Poverty Incidence Per Region for 2012 and 2015**

## **Recommendation and Conclusion**