Final Assignment

Ananya Chatterjee, UCO: 459203

Solution 1

Probability Density Function:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[\frac{-(ln(x)-\mu)^2}{2\sigma^2}\right]$$

The likelihood function can be written as:

$$L(\mu|x) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} exp\left[\frac{-(ln(x_i)-\mu)^2}{2\sigma^2}\right]$$

$$= \frac{1}{(\sigma^2 2\pi)^{\frac{n}{2}}} exp\left[\frac{\sum_{i=1}^{n} -(ln(x_i)-\mu)^2}{2\sigma^2}\right]$$

Log likelihood function is:

$$l(\mu|x) = \frac{-\sum_{i=1}^{n} -(ln(x_i)-\mu)^2}{2\sigma^2} - \frac{nln2\pi}{2} - nln\sigma$$

Score function:

$$S(\mu) = \frac{dl(\mu|x)}{dt} = \frac{\sum_{i=1}^{n} -(ln(x_i) - n\mu)}{\sigma^2}$$

Now we know that $S(\mu) = 0$ for maximal likelihood estimate.

Therefore, $\sum_{i=1}^{n} -(ln(x_i) - n\mu) = 0$

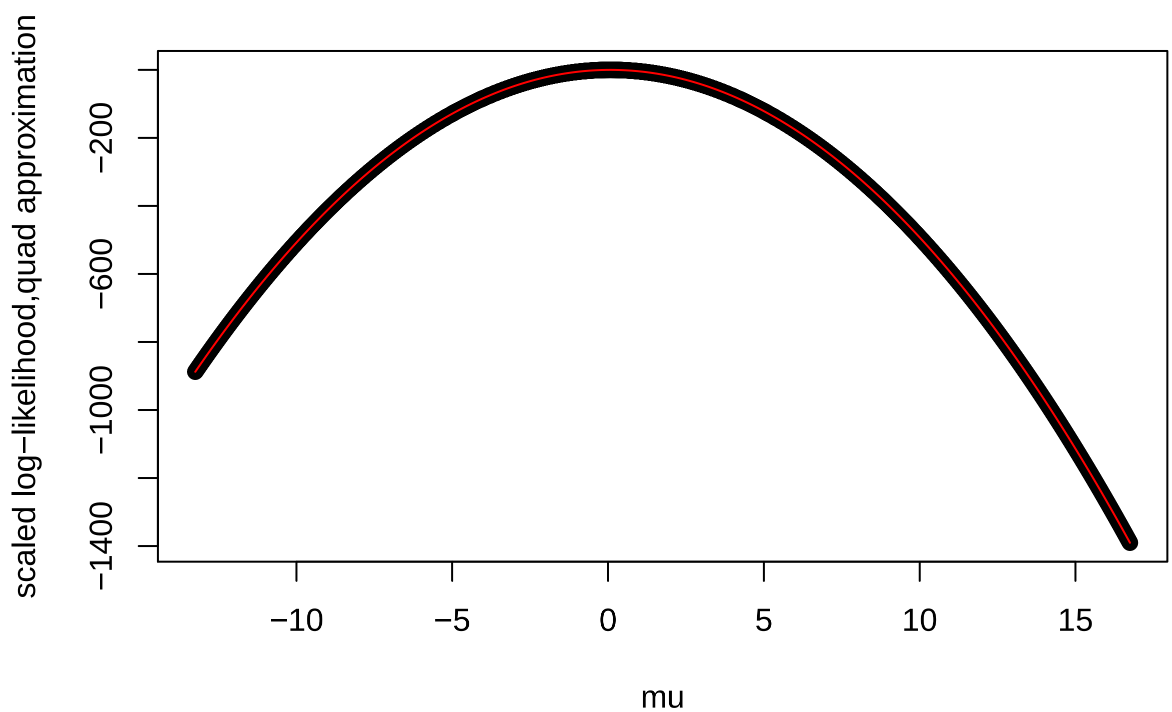
$$\mu = \frac{\sum_{i=1}^{n} ln(x_i)}{n}$$

Fisher Information:

$$I(\hat{\mu}) = -\frac{d^2 l(\mu|x)}{du^2} = \frac{n}{\sigma^2}$$

```
x = c(4.856, 0.487, 0.580, 0.839, 0.721, 2.416, 0.715, 0.361, 0.703, 5.829)
sigma = 1
n = length(x)
loglikelihood = function(mu)
{
  l = -((sum((log(x)-mu)^2))/(2*(sigma^2)))-((n*log(2*pi))/2)-(n*log(sigma))
  return (l)
}
mu1 = sum(x)/n
mu2 = seq(mu1-15,mu1+15,by=0.01)
l_mu_x = matrix(0,length(mu2),1)

#Log-likelihood over the range mu2
for(i in 1:length(mu2))
{
  l_mu_x[i] = loglikelihood(mu2[i])
}
max_l = max(l_mu_x)

#Scaled log-likelihood
```

```
scaled_l = l_mu_x - max_l

plot(mu2,scaled_l,col="black",xlab="mu",ylab="scaled log-likelihood,quad approximation")

#Quadratic Approximation
quad_approx = function(fisher,mu_est,mu)
{
  q = -(0.5*fisher*((mu-mu_est)^2))
  return (q)
}
fisher = n/((sigma^2))

mu_est = (sum(log(x)))/n
quad_approx_scaled_l = matrix(0,length(mu2),1)

#Quadratic approximation over mu2
for(i in 1:length(mu2))
{
  quad_approx_scaled_l[i] = quad_approx(fisher,mu_est,mu2[i])
}
lines(mu2,quad_approx_scaled_l,col="red")
```



Solution 2

H_0 : mean of skull length is equal to 177.568 mm

H_1 : mean of skull length is not equal to 177.568 mm

Wald Test Statistic Tw:

$Tw = (\mu_h - \mu_0)\sqrt{I(\mu_h)}$ follows $N(0,1)$

$\mu_h = \frac{\sum_{i=1}^{n} x_i}{n} = mean(x)$ for $N(\mu, \sigma^2)$

$\mu_0 = 177.568$

$I(\mu_h) = \frac{n}{\sigma^2}$

```
mydata = read.table("one-sample-mean-skull-mf.txt",header = TRUE)
skull_male=subset(mydata,sex=="m",select=skull.L)
mu0 = 177.568

#Remove NA entrie(s)
skull_male = skull_male[!is.na(skull_male)]

n=length(skull_male)
mu_hat = mean(skull_male)
variance = var(skull_male)
Tw = (mu_hat - mu0)*(sqrt(n/variance))
print(Tw)
```

```
## [1] 10.33382
```

```
#For N(0,1)
print(qnorm(0.025))
```

```
## [1] -1.959964
```

```
print(qnorm(0.975))
```

```
## [1] 1.959964
```

```
#Critical region is (-infinity,qnorm(0.025))U(qnorm(0.975),infinity) = (-infinity,-1.96)U(1.96,infinity
#Tw=10.33382 belongs to it, we reject NULL Hypothesis
```

Likelihood Ratio Test Statistics ULR

$ULR = -2 * [l(\mu_0|x) - l(\hat{\mu}|x)]$

```
sigma = sd(skull_male)
loglikelihood = function(mu)
{
  l = -((n/2)*(log(2*pi)))-((n/2)*(log(sigma^2)))-( (sum((skull_male-mu)^2))/(2*sigma^2) )
  return (l)
}

#Log-likelihood at mu = mu0
loglikelihood_mu0 = loglikelihood(mu0)
print(loglikelihood_mu0)
```

```
## [1] -762.6142
```

```r
#Log-likelihood at mu = mu_hat
loglikelihood_mu_hat = loglikelihood(mu_hat)
print(loglikelihood_mu_hat)
```

```
## [1] -709.2203
```

```r
ULR = -2*(loglikelihood_mu0 - loglikelihood_mu_hat)
print(ULR)
```

```
## [1] 106.7877
```

```r
#Chi square lower limit
print(qchisq(0.95,1))
```

```
## [1] 3.841459
```

```r
#ULR=106.7877 belongs to (qchisq(0.95,1),infinity) = (3.841459,infinity)
#We reject the NULL Hypothesis
```

Wald Empirical Confidence Interval

$$(d_w, h_w) = (\hat{\mu} - \mu_{\alpha/2} * \frac{1}{\sqrt{I(\hat{\mu}}}, \hat{\mu} + \mu_{\alpha/2} * \frac{1}{\sqrt{I(\hat{\mu}}})$$

```r
mu_alpha = qnorm(0.975)
#Lower limit of confidence interval
dw = mu_hat - (mu_alpha*(sqrt(variance/n)))
#Upper limit of confidence interval
hw = mu_hat + (mu_alpha*(sqrt(variance/n)))
print(dw)
```

```
## [1] 181.1893
```

```r
print(hw)
```

```
## [1] 182.8845
```

```r
#Wald Empirical Confidence Interval is (dw,hw) = (181.1893,182.8845).
```

Likelihood Ratio Confidence Interval

$$CS_{1-\alpha} = \mu : -2 * [l(\mu_0|x) - l(\hat{\mu}|x) - {\psi_1}^2(\alpha)] < 0$$

```r
mu = seq(180, 183, by = 0.01)
mat = matrix(0,length(mu),1)
for(i in 1:length(mu))
{
  mat[i] = -2 * (loglikelihood(mu[i]) - loglikelihood(mu_hat)) - qchisq(0.95,1)
}
```

```r
f1 = function(mu){
  l = -2 * (loglikelihood(mu) - loglikelihood(mu_hat)) - qchisq(0.95,1)
}
#Compute lower limit of confidence interval
LL = uniroot(f1,c(180, mu_hat))$root
print(LL)
```

```
## [1] 181.1893
```

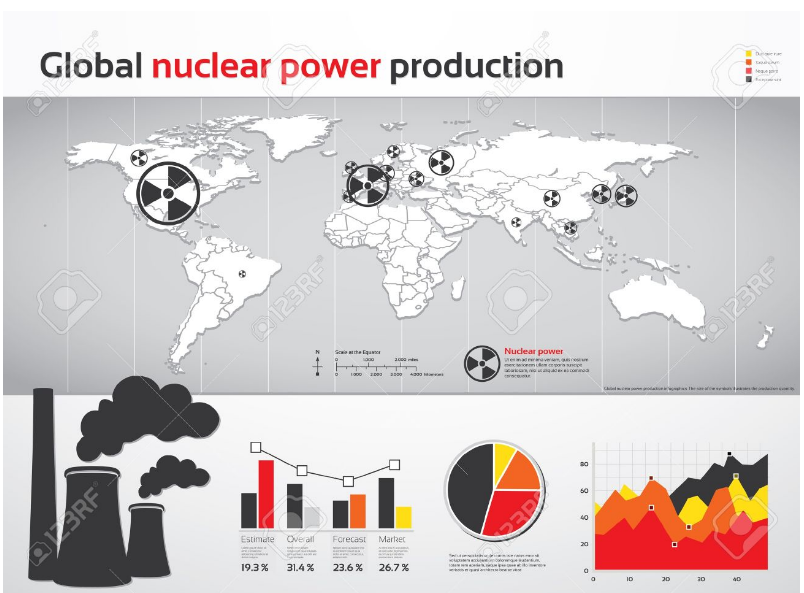
```r
#Compute upper limit of confidence interval
UL = uniroot(f1,c(mu_hat, 185))$root
print(UL)
```

```
## [1] 182.8845
```

```r
#Likelihood ratio confidence interval = (181.1893,182.8845)
```

Solution 3

```r
setwd("/home/ananya/ICT-workbench/Statistics/MV013_Final")
library(ggplot2)
library(png)
#Read and Display PNG Image
DisplayPNGImage = function(path, aspect){
require(’png’)
img = readPNG(path, native=T)
res = dim(img)[1:2]
plot(1, 1, xlim=c(1,res[1]), ylim=c(1,res[2]), type=’n’, xaxs=’i’, yaxs=’i’, xaxt=’n’,yaxt=’n’,xlab=’’,
rasterImage(img, 1, 1, res[1], res[2])
}

#Example of First Good Graph
```

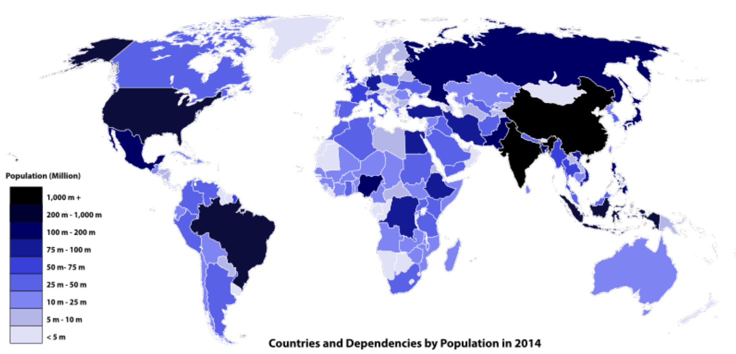


1. Large amount of information represented in understandable form.
2. Eye catching colors with good contrast
3. Easy to understand

```
#Example of Second Good Graph
```

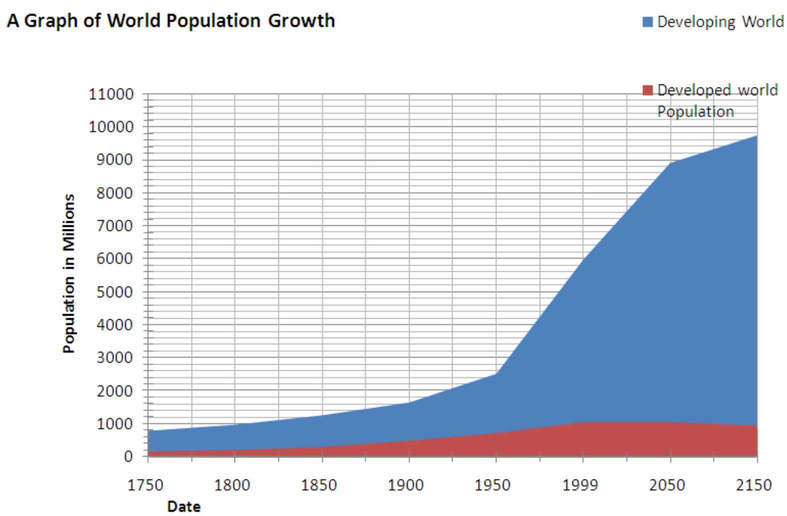
1. Information given can be understood by all.
2. Eye catching colors with good contrast
3. Covers large amount of information in one picture.

```
#Example of Third Good Graph
```



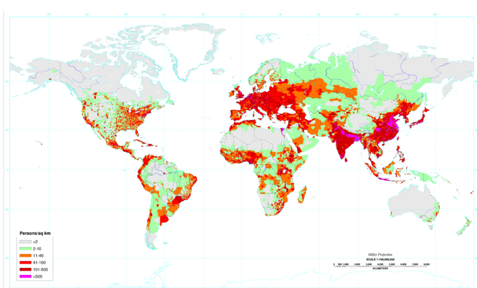
1. Use of soothing colors
2. Easy to understand and self explanatory.
3. Legend given is easy to understand.

```
#Example of First Bad Graph
```



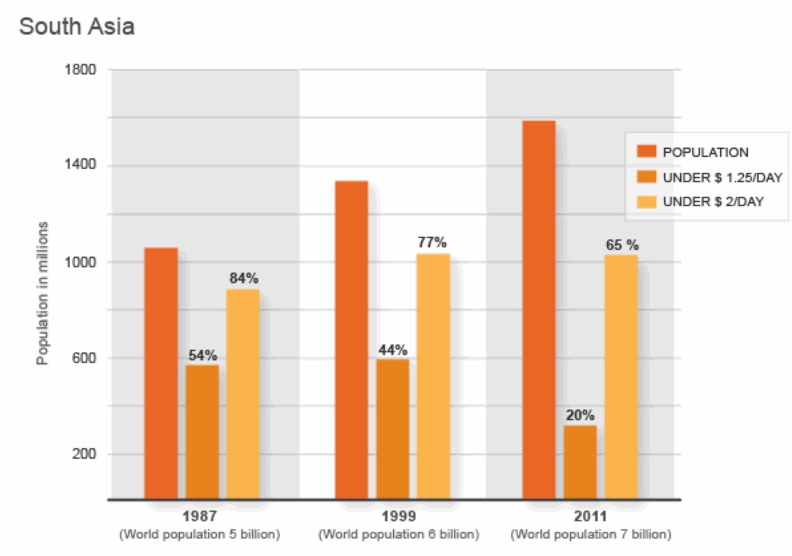
1. Area Chart is not clear about population growth
2. Yearwise population growth could have been shown by drawing a seperate graph.

```
#Example of Second Bad Graph
```



1. Legend provided is not clear on what it is measuring.
2. The colours used are not contrasting and creates confusion

```
#Example of Third Bad Graph
```

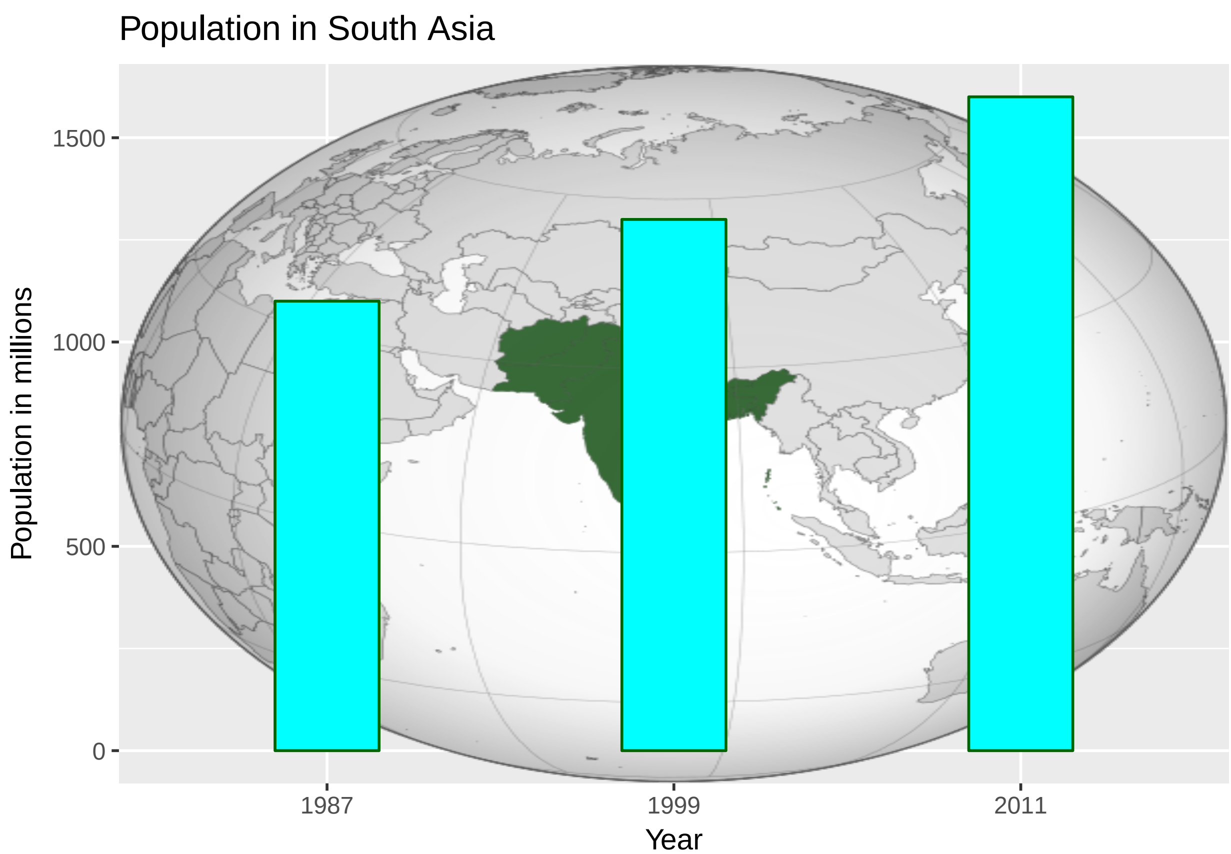


1. Seperate pie charts could have been drawn for income calculation
2. The colours used are not contrasting enough for clear differentiation

Improvement of a Bad Example

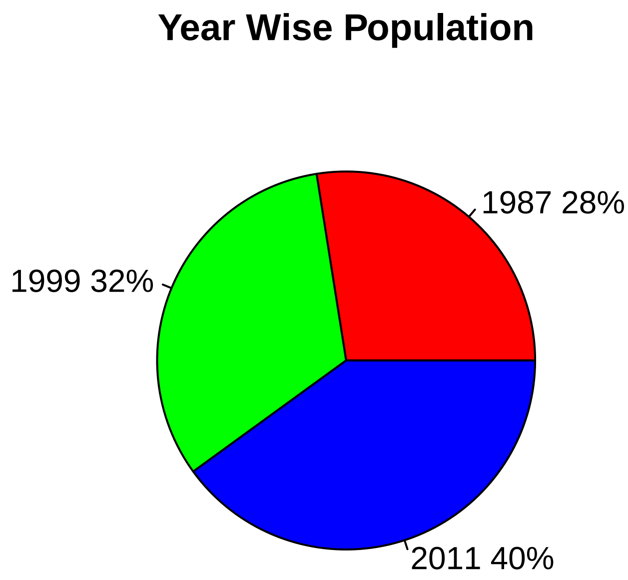
Let us the numbers from the Bad Graph 3 and draw the graph with improvements.

```
library(grid)
library(ggplot2)
library(plotrix)
library(gridExtra)
require('png')
image <- readPNG("South_Asia.png")
mydata <- data.frame(

Year = factor(c("1987","1999", "2011"),
levels=c("1987","1999", "2011")),
Population = c(1100, 1300, 1600)
)

ggplot(data = mydata, aes(x = Year, y = Population)) +
annotation_custom(rasterGrob(image,
width = unit(1,"npc"),
height = unit(1,"npc")),
-Inf, Inf, -Inf, Inf) +
geom_bar(stat = "identity", position = position_dodge(),width=0.3, fill="cyan", color="darkgreen") +
xlab("Year") + ylab("Population in millions") +
ggtitle("Population in South Asia")
```

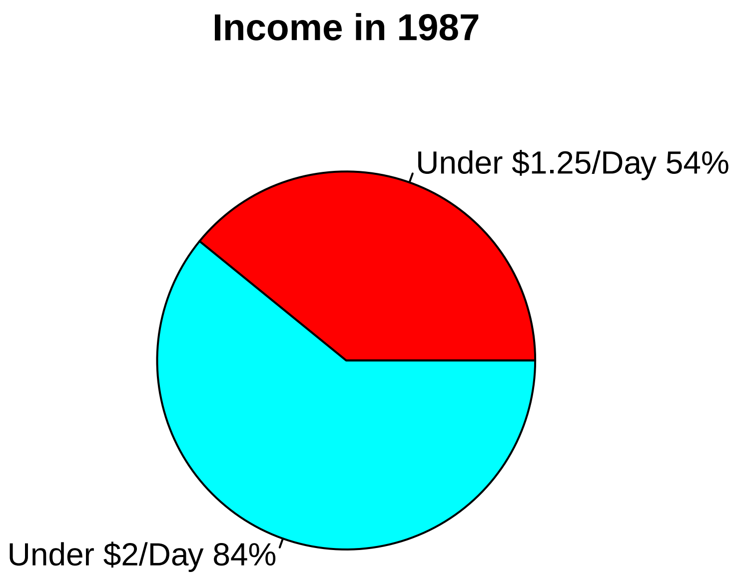


Pie Charts have been made for repesenting comparisons of Daily Population Income of three years

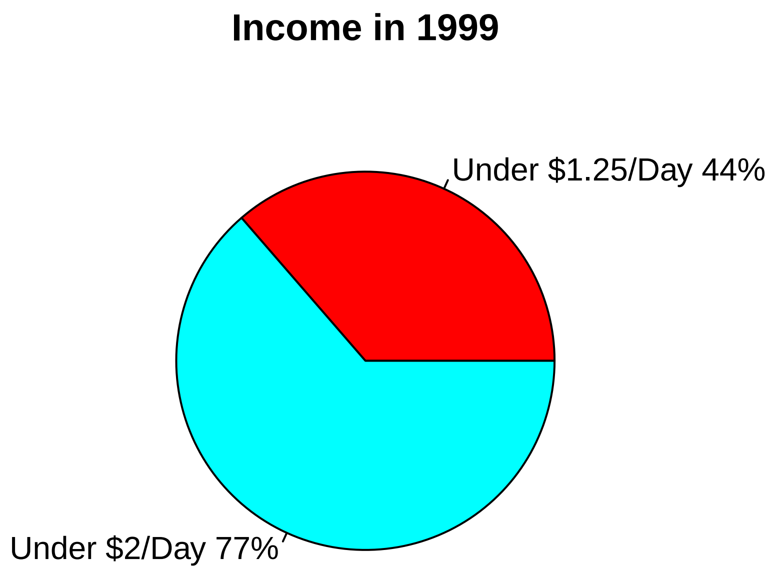
```
slices <- c(1100, 1300, 1600)
lbls <- c("1987","1999","2011")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct)
lbls <- paste(lbls,"%",sep="")
pie(slices,labels = lbls, col=rainbow(length(lbls)),
main="Year Wise Population")
```



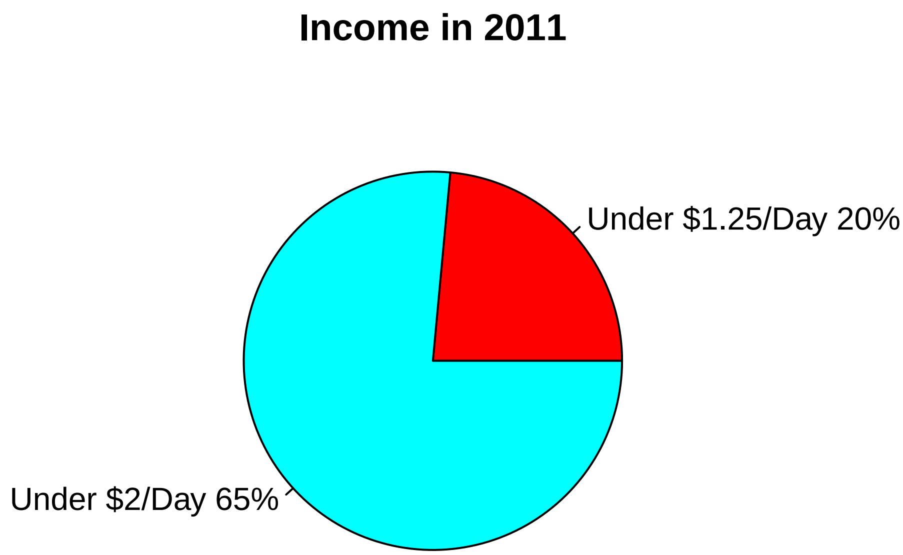

```
slices <- c(54, 84)
lbls <- c("Under $1.25/Day", "Under $2/Day")
lbls_1 <- paste(lbls, slices)
lbls_1 <- paste(lbls_1,"%",sep="")
p <- pie(slices, labels = lbls_1, col=rainbow(length(lbls)),
main="Income in 1987")
```




```
slices <- c(44, 77)
lbls_2 <- paste(lbls, slices)
lbls_2 <- paste(lbls_2,"%",sep="")
p <- pie(slices, labels = lbls_2, col=rainbow(length(lbls)),
main="Income in 1999")
```




```
slices <- c(20, 65)
lbls_3 <- paste(lbls, slices)
lbls_3 <- paste(lbls_3,"%",sep="")
p <- pie(slices, labels = lbls_3, col=rainbow(length(lbls)),
main="Income in 2011")
```


Income in 2011
Under $1.25/Day 20%
Under $2/Day 65%