```r
rm(list = ls())
setwd("G:\\math\\640")
options(scipen=999)
source("multiplot.R")
require(VGAM, quietly=T); require(invgamma, quietly=T); require(mvtnorm, quietly=T);
```

```
## Warning: package 'VGAM' was built under R version 3.5.2
```

```
## Warning: package 'invgamma' was built under R version 3.5.2
```

```
## Warning: package 'mvtnorm' was built under R version 3.5.2
```

```r
require(ggplot2, quietly=T); require(RColorBrewer, quietly=T); require(zoo, quietly=T)
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
wind <- read.table("wind.txt",header = F)
```

$$f(w_i) = \frac{w_i}{\theta^2}\,\exp\left(-\frac{w_i^2}{2\theta^2}\right)$$

$$\begin{aligned}\mathcal{L}(W|\theta^2) &\propto \prod_{i=1}^{n}\frac{1}{\theta^2}\,\exp\left(-\frac{w_i^2}{2\theta^2}\right)\\ &= \left(\theta^2\right)^{-n}\,\exp\left(-\frac{1}{2\theta^2}\sum_{i=1}^{n}w_i^2\right)\end{aligned}$$

A conjugate prior for θ^2 is the inverse gamma

$$\pi\left(\theta^2\right) \propto \left(\theta^2\right)^{-(\alpha_0+1)}\,\exp\left(-\frac{\beta_0}{\theta^2}\right).$$

And if we let $\alpha_0 = \beta_0 = 0$, we get an improper, non-informative flat prior

$$\pi\left(\theta^2\right) \propto \left(\theta^2\right)^{-1}.$$

Which leads to the posterior

$$P(\theta^2|W) \propto \left(\theta^2\right)^{n+1}\,\exp\left(-\frac{1}{2\theta^2}\sum_{i=1}^{n}w_i^2\right).$$

And

$$\theta^2|W \sim IG\left(n,\ \frac{1}{2}\sum_{i=1}^{n} w_i^2\right).$$

Or we could transform $\theta^2 = \mu^{-2}$

$$\begin{aligned}\mathcal{L}(W|\mu^{-2}) &\propto \prod_{i=1}^{n} \mu^2 \ \exp\left(-\mu^2 \frac{w_i^2}{2}\right) \\ &= \left(\mu^2\right)^n \ \exp\left(-\frac{\mu^2}{2}\sum_{i=1}^{n} w_i^2\right)\end{aligned}$$

$$\ell(W|\mu^{-2}) \propto n \log\left(\mu^2\right) - \frac{\mu^2}{2}\sum_{i=1}^{n} w_i^2$$

$$\frac{\partial \ell}{\partial\left(\mu^2\right)} \propto \frac{n}{\mu^2} - \frac{1}{2}\sum_{i=1}^{n} w_i^2$$

$$\frac{\partial^2 \ell}{\partial\left(\mu^2\right)^2} \propto -\frac{n}{\left(\mu^2\right)^2}$$

$$\begin{aligned}-E\left[\frac{\partial^2 \ell}{\partial\left(\mu^2\right)^2}\right] &= \left(\mu^2\right)^{-2} \\ \left[J\left(\mu^2\right)\right]^{\frac{1}{2}} &= \left(\mu^2\right)^{-1}\end{aligned}$$

$$P\left(\mu^2|W\right) \propto \left(\mu^2\right)^{n-1} \ \exp\left(-\frac{\mu^2}{2}\sum_{i=1}^{n} w_i^2\right)$$

$$\mu^2|W \sim Gamma\left(n,\ \frac{1}{2}\sum_{i=1}^{n} w_i^2\right)$$

```
n <- nrow(wind)
Nsim <- 2000
set.seed(17)
thetatheta <- rinvgamma(Nsim , n+1 , (1/2)* sum(wind[,1]^2)   )
quantile(thetatheta , probs = c(0.025, 0.5, 0.975))
```

```
##     2.5%      50%    97.5%
## 47.54592 55.42074 65.45862
```

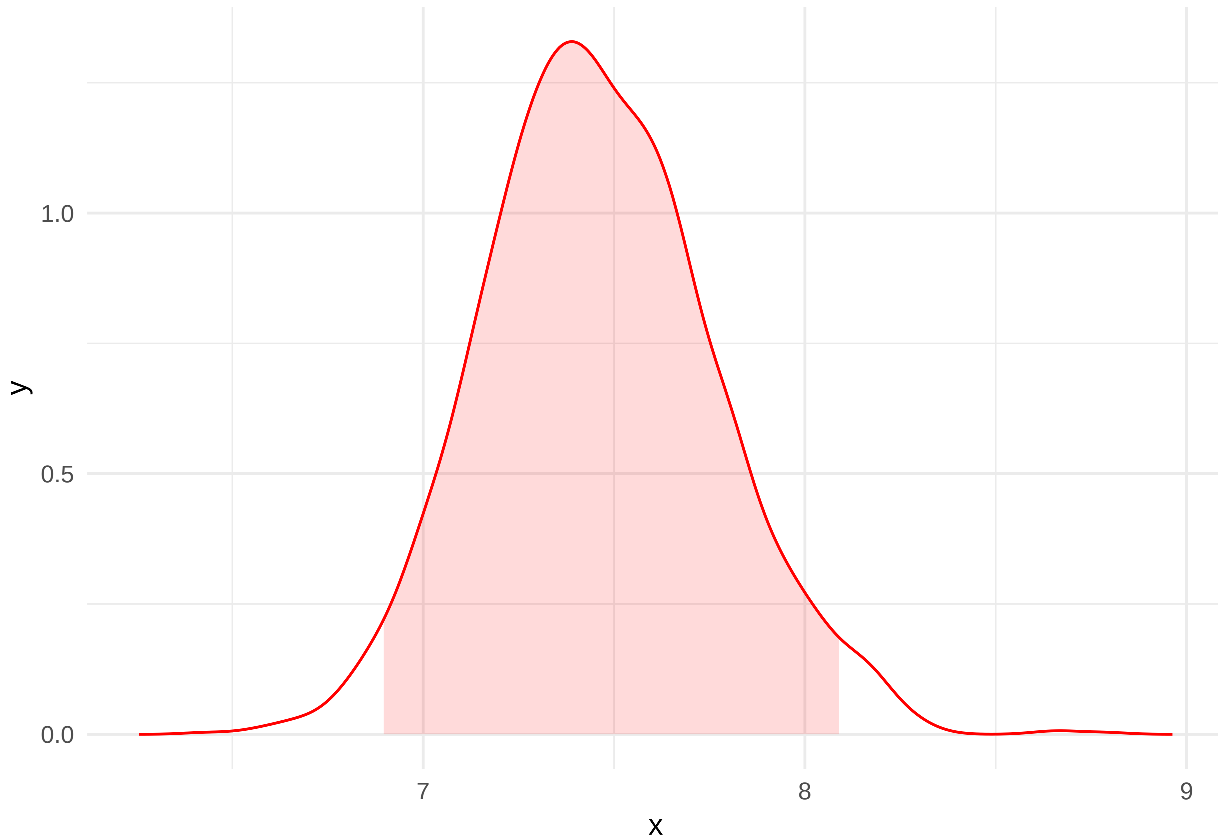
```
#mode
theta <- sqrt( thetatheta  )
summary(theta)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.431   7.255   7.445   7.459   7.654   8.787
```

```r
quantile(theta, probs = c(0.025, 0.5, 0.975))
```

```
##     2.5%      50%    97.5%
## 6.895355 7.444511 8.090650
```

```r
qu <- quantile(theta , probs = c(0.5, 0.025, 0.975))[2:3]
d <- data.frame( density(theta)[[1]], round(density(theta)[[2]],6) )
colnames(d) = c("x" , "y" ); d$area <- d[,1] > qu[1] & d[,1] < qu[2]

ggplot( data = d , aes(x=x , y=y)    ) + geom_line( col="red"  ) +
    geom_ribbon(data = d[which(d$area == T),], aes(x, ymin=0 , ymax=y ), fill="red", alpha = .15) +
    theme(legend.position="none", axis.title.x=element_blank(), axis.title.y=element_blank(),
         plot.title = element_text(hjust = 0.5)   )+ theme_minimal()
```




```r
rmean <- theta * sqrt(  pi / 2 )

summary(rmean )
```

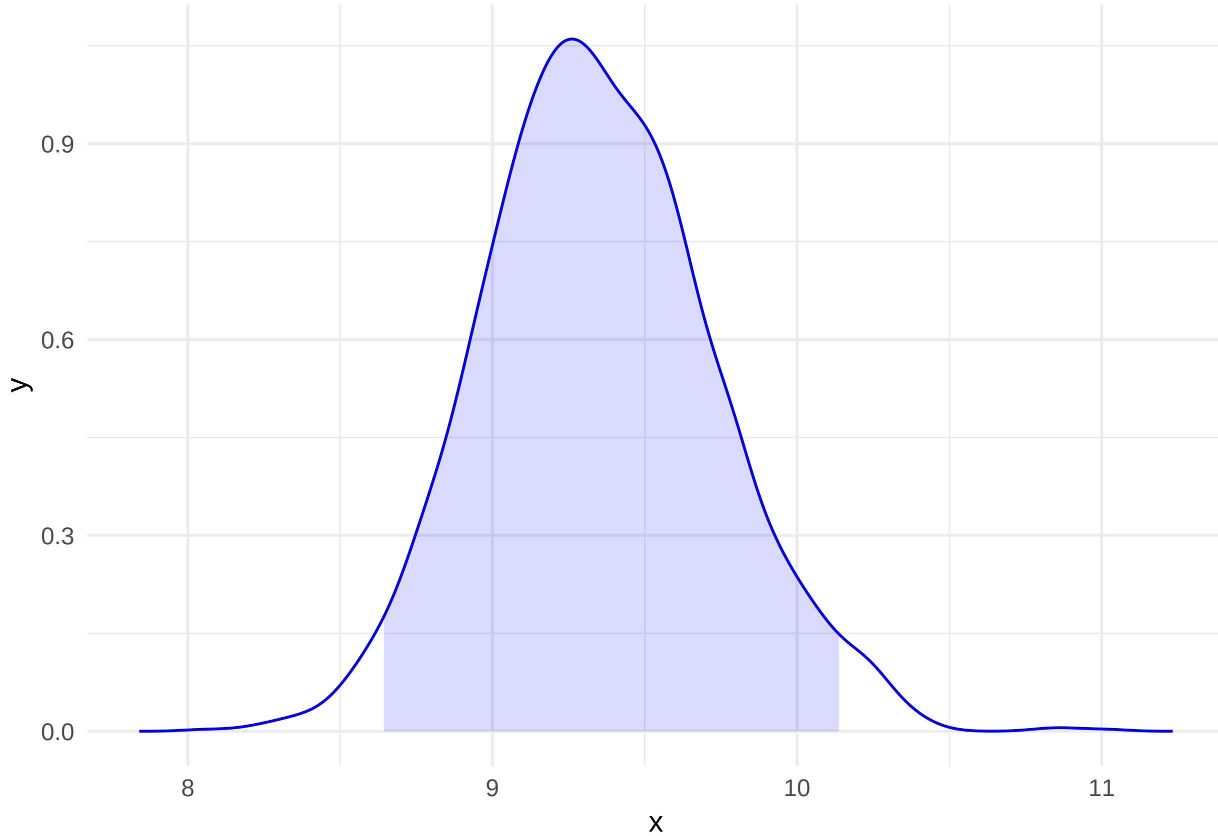
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   8.060   9.093   9.330   9.348   9.593  11.013
```

```r
quantile(rmean, probs = c(0.025, 0.5, 0.975))
```

```
##      2.5%       50%     97.5%
##  8.642046  9.330311 10.140126
```

```r
qu <- quantile(rmean, probs = c(0.5, 0.025, 0.975))[2:3]
d <- data.frame( density(rmean)[[1]], round(density(rmean)[[2]],6) )
colnames(d) = c("x" , "y" ); d$area <- d[,1] > qu[1] & d[,1] < qu[2]

ggplot( data = d , aes(x=x , y=y)    ) + geom_line( col="blue"  ) +
```

```
    geom_ribbon(data = d[which(d$area == T),], aes(x, ymin=0 , ymax=y ), fill="blue", alpha = .15) +
    theme(legend.position="none", axis.title.x=element_blank(), axis.title.y=element_blank(),
         plot.title = element_text(hjust = 0.5)    )+ theme_minimal()
```




```
rmedian <- theta * sqrt(  2 * log(2) )
summary(rmedian )
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   7.572   8.542   8.765   8.782   9.012  10.346
```

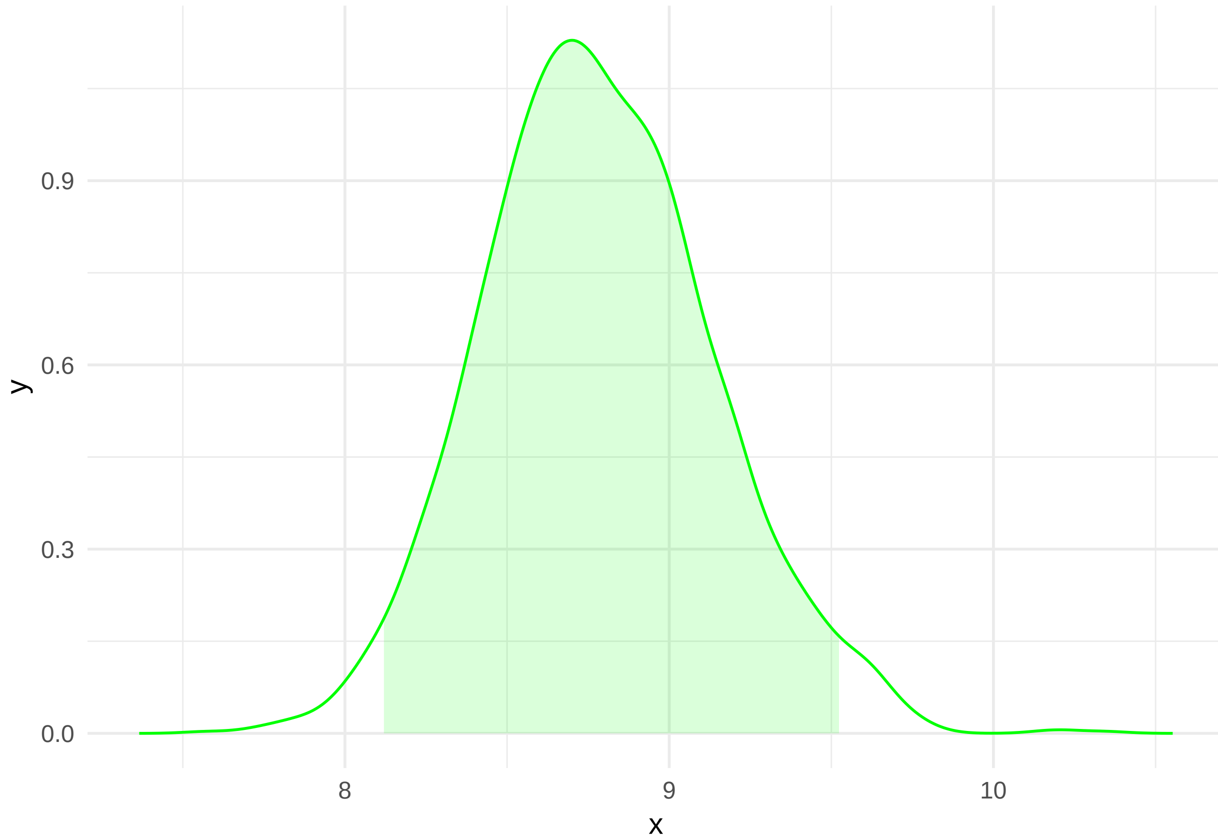
```
summary(rmedian )
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   7.572   8.542   8.765   8.782   9.012  10.346
```

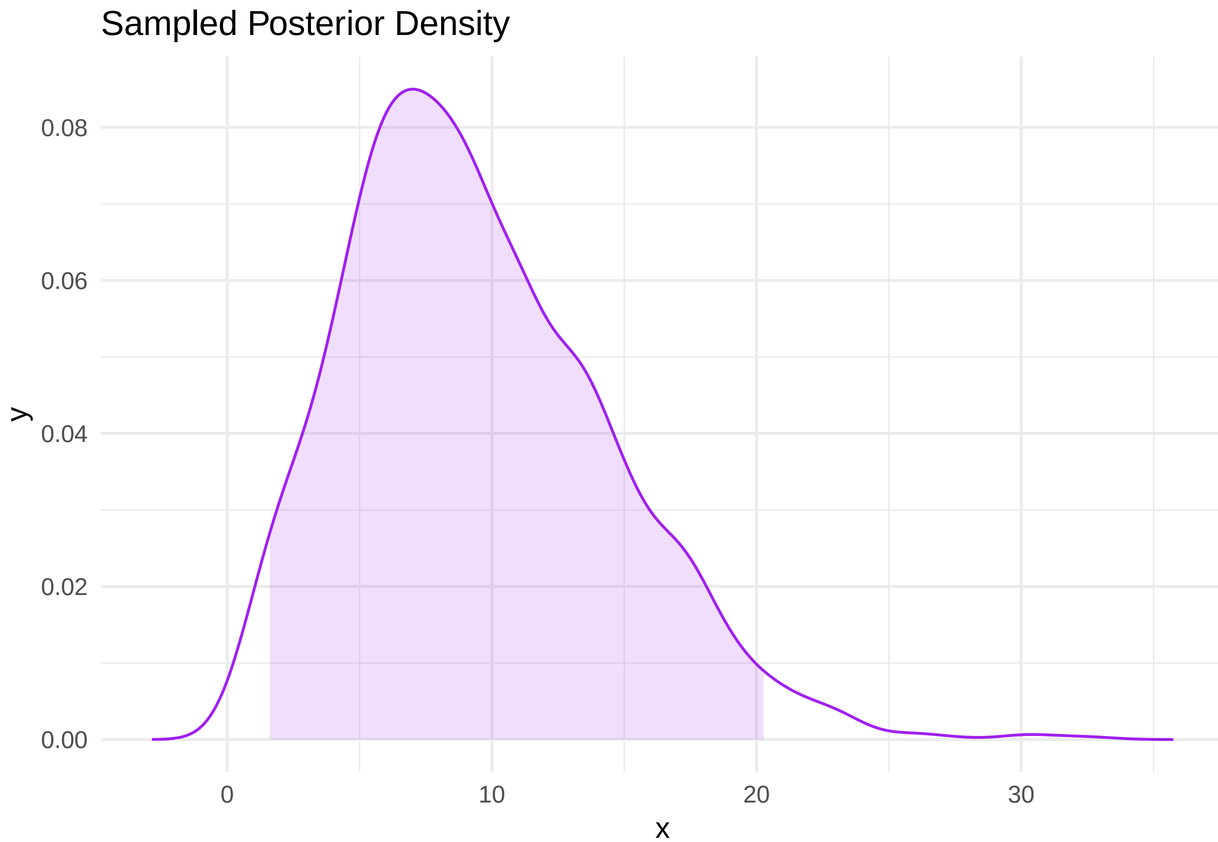
```
quantile(rmedian, probs = c(0.025, 0.5, 0.975))
```

```
##     2.5%      50%    97.5%
## 8.118660 8.765242 9.526012
```

```
qu <- quantile(rmedian, probs = c(0.5, 0.025, 0.975))[2:3]
d <- data.frame( density(rmedian)[[1]], round(density(rmedian)[[2]],6) )
colnames(d) = c("x" , "y" ); d$area <- d[,1] > qu[1] & d[,1] < qu[2]

ggplot( data = d , aes(x=x , y=y)    ) + geom_line( col="green"  ) +
    geom_ribbon(data = d[which(d$area == T),], aes(x, ymin=0 , ymax=y ), fill="green", alpha = .15) +
    theme(legend.position="none", axis.title.x=element_blank(), axis.title.y=element_blank(),
         plot.title = element_text(hjust = 0.5)    )+ theme_minimal()
```




```
PPD <- thetatheta *0
set.seed(17)
for( i in 1:Nsim){  PPD[i] <- rrayleigh(1,  sqrt(thetatheta[i])) }

qu <- quantile(PPD, probs = c(0.5, 0.025, 0.975))[2:3]
d <- data.frame( density(PPD)[[1]], round(density(PPD)[[2]],6) )
colnames(d) = c("x" , "y" ); d$area <- d[,1] > qu[1] & d[,1] < qu[2]

ggplot( data = d , aes(x=x , y=y)    ) + geom_line( col="purple"  ) +
    geom_ribbon(data = d[which(d$area == T),], aes(x, ymin=0 , ymax=y ), fill="purple", alpha = .15) +
    theme(legend.position="none", axis.title.x=element_blank(), axis.title.y=element_blank(),
         plot.title = element_text(hjust = 0.5)   )+ theme_minimal()+
   ggtitle("Sampled Posterior Density")
```

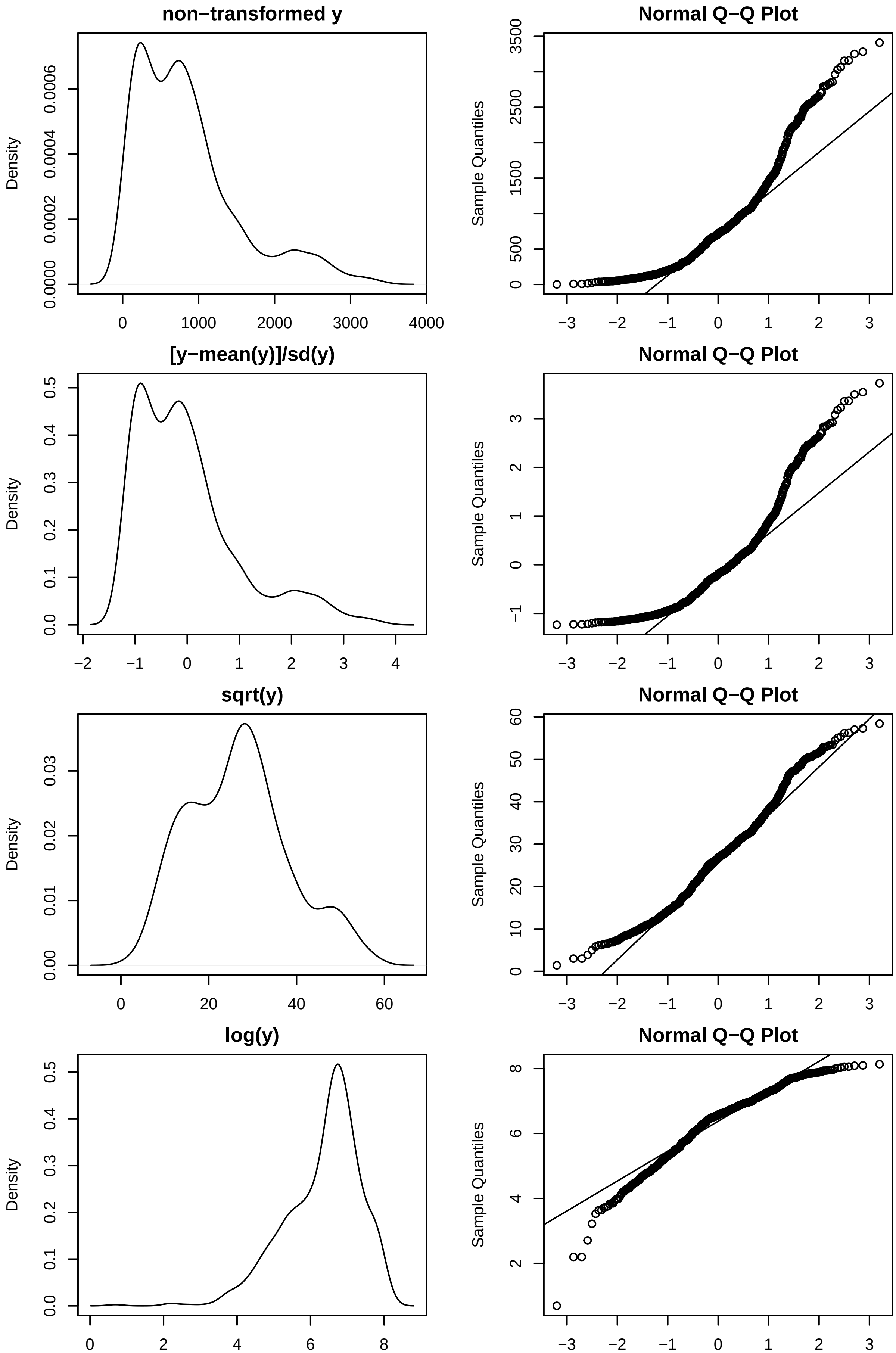



```
sum(1*(PPD > 15))/length(PPD)
```

```
## [1] 0.14
```

The first model will be the casual users y_c. We first check to see if we should transform the outcome variable.

```
cb <- read.table("day.txt",header = T)

par(mfrow=c(4,2) ,  mar=c(2.1,4.1,2.1,2.1) )
ty <-  (  cb$casual ); plot(density(ty), main="non-transformed y"); qqnorm(ty); qqline(ty )
ty <- scale(cb$casual) ; plot(density(ty), main="[y-mean(y)]/sd(y)"); qqnorm(ty); qqline(ty )
ty <- sqrt(  cb$casual ); plot(density(ty), main="sqrt(y)"); qqnorm(ty); qqline(ty )
ty <- log(  cb$casual ); plot(density(ty), main="log(y)"); qqnorm(ty); qqline(ty )
```


non−transformed y
Density
0.0000
0.0002
0.0004
0.0006
0
1000
2000
3000
4000
Normal Q−Q Plot
Sample Quantiles
0
500
1500
2500
3500
[y−mean(y)]/sd(y)
Density
0.0
0.1
0.2
0.3
0.4
0.5
−2
−1
0
1
2
3
4
Normal Q−Q Plot
Sample Quantiles
−1
0
1
2
3
sqrt(y)
Density
0.00
0.01
0.02
0.03
0
20
40
60
Normal Q−Q Plot
Sample Quantiles
0
10
20
30
40
50
60
log(y)
Density
0.0
0.1
0.2
0.3
0.4
0.5
0
2
4
6
8
Normal Q−Q Plot
Sample Quantiles
2
4
6
8
−3
−2
−1
0
1
2
3

The first row of plots shows that y_c should probably be transformed to ensure normality. Of the three transformation choices presented here, it appears $\sqrt{y_c}$ provides a reasonable assumption of normality. Using example I.22 we find the following results.

Now assume the transformed variable y_c' has the distribution

$$y_c' \sim N(X\boldsymbol{\beta}, \sigma^2 I_n)$$

The likelihood is then

$$\begin{aligned}\mathcal{L}\left(y|X,\boldsymbol{\beta},\sigma^2\right) &\propto \left|\sigma^2 I_n\right|^{-\frac{1}{2}} \ \exp\left[-\frac{1}{2}(Y - X\boldsymbol{\beta})^T \left(\sigma^2 I_n\right)(y_c' - X\boldsymbol{\beta})\right] \\ &= \left(\sigma^2\right)^{-\frac{n}{2}} \ \exp\left[-\frac{1}{2\sigma^2}(y_c' - X\boldsymbol{\beta})^T (y_c' - X\boldsymbol{\beta})\right]\end{aligned}$$

With the joint prior

$$\pi\left(\boldsymbol{\beta},\sigma^2\right) \propto N(X\boldsymbol{\beta}, \sigma^2 I_n)$$

which gives a joint posterior of

$$P\left(\boldsymbol{\beta},\sigma^2|y,X\right) \propto \left(\sigma^2\right)^{-\frac{n}{2}-1} \ \exp\left[-\frac{1}{2\sigma^2}(y_c' - X\boldsymbol{\beta})^T (y_c' - X\boldsymbol{\beta})\right]$$

We find

$$\boldsymbol{\beta}|\sigma^2 \sim N\left[\hat{\boldsymbol{\beta}}, \sigma^2\left[X^T X\right)^{-1}\right]$$

and

$$\sigma^2|X \sim IG\left[\frac{n-k}{2}, \frac{1}{2}\left(y_c' - X\hat{\boldsymbol{\beta}}\right)^T \left(y_c' - X\hat{\boldsymbol{\beta}}\right)\right].$$

Now we can perform the analysis.

```
n <- nrow(cb )
Nsim <- 20000
y1 <- sqrt( cb[,1] )     #casual
X <- as.matrix( cbind(1, cb[,-(1:2)] ) )
colnames(X) <- c("(Intercept)",colnames(cb[,-(1:2)] ))
p <- ncol(X)

bhat <-  (solve(t(X)%*%X)%*%(t(X)%*%y1))
#fit <- ( lm( y1~X[,-1]  ));
```

```r
SSY <- SSYc <- t(y1 - X%*%bhat)%*%(y1 - X%*%bhat); #anova(fit)
#vcov( fit )
rbetas <- matrix(0, nrow = Nsim, ncol = p)
XtXi    <- solve(t(X)%*%X)

set.seed(13)
rsig    <- rinvgamma(Nsim , (n-p)/2, (1/2)*SSY)
for(i in 1:Nsim){
    CovX    <- rsig[i]*XtXi
    rbetas[i,] <- c(rmvnorm(1, mean = bhat, sigma = CovX ) )
}

ycrbetas <- rbetas
rbMat   <- apply(rbetas, 2, quantile, probs = c(0.5, 0.025, 0.975))
colnames(rbMat) <- colnames(X) #coef(fit)
rbMat <- as.data.frame( round(t(rbMat), 4) )
rbMat$zero <- rbMat[,2] < 0 & rbMat[,3] > 0
rbMat
```

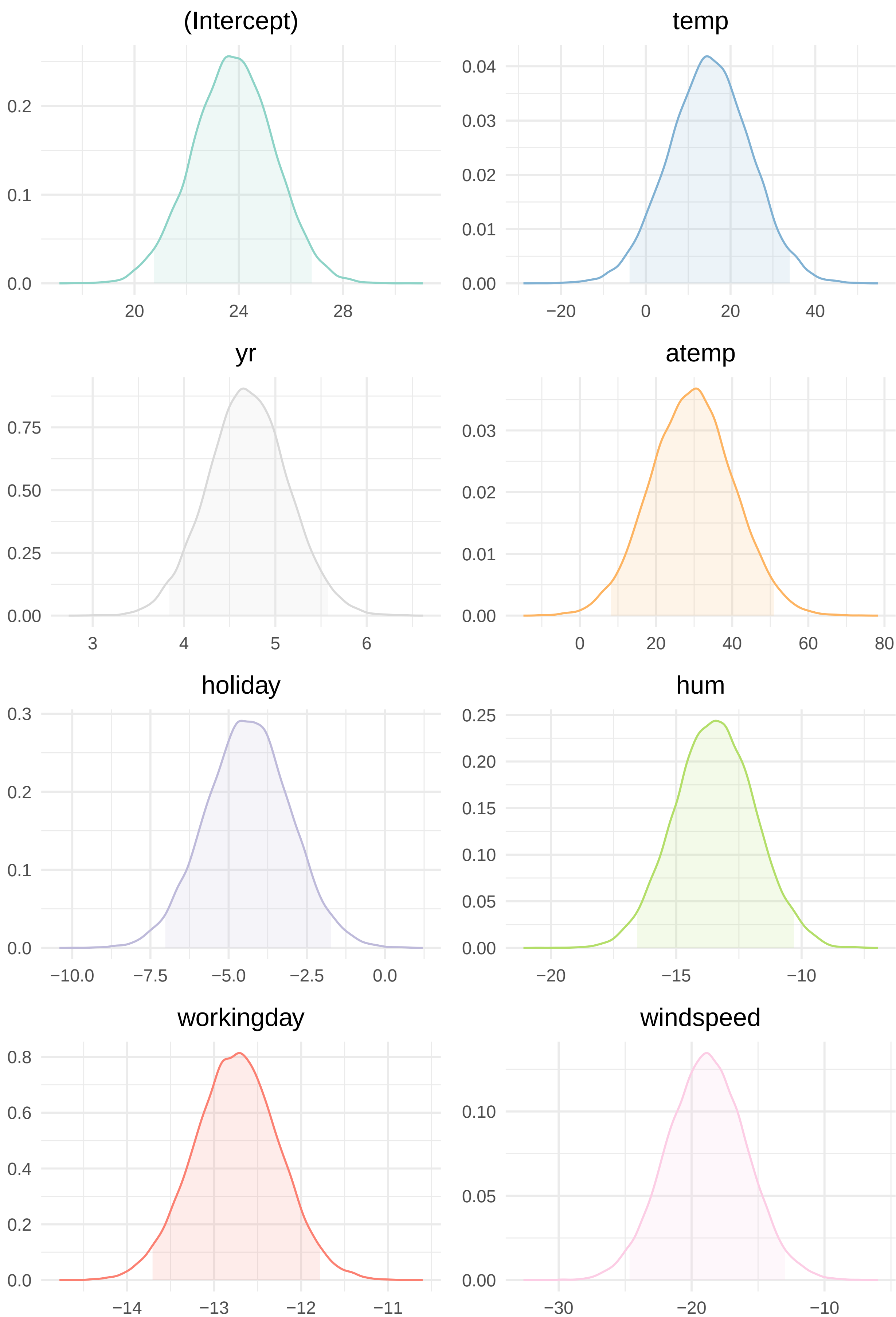
```
##                   50%     2.5%    97.5%  zero
## (Intercept)  23.8007  20.7258  26.8012 FALSE
## yr            4.6969   3.8357   5.5787 FALSE
## holiday      -4.3741  -7.0237  -1.7099 FALSE
## workingday  -12.7406 -13.7104 -11.7732 FALSE
## temp         15.0892  -3.8413  34.1364  TRUE
## atemp        29.5032   8.0181  50.9557 FALSE
## hum         -13.4621 -16.5841 -10.2962 FALSE
## windspeed   -18.8572 -24.7187 -12.9516 FALSE
```

```r
rbsum <- ( apply( rbetas , 2 , summary) )
colnames(rbsum )    <- colnames(X)
 rbsum <- t( rbsum ); rbsum
```

```
##                   Min.    1st Qu.     Median       Mean    3rd Qu.
## (Intercept)  17.697267  22.750111  23.800728  23.797447  24.847835
## yr            2.901046   4.405509   4.696941   4.700773   4.993859
## holiday      -9.908031  -5.277078  -4.374120  -4.375107  -3.475345
## workingday  -14.596365 -13.069972 -12.740636 -12.741467 -12.413182
## temp        -25.288433   8.642124  15.089204  15.109417  21.613841
## atemp       -10.765097  22.075386  29.503219  29.489856  36.733933
## hum         -20.499974 -14.550989 -13.462148 -13.461465 -12.371242
## windspeed   -31.536690 -20.900125 -18.857223 -18.860783 -16.844467
##                    Max.
## (Intercept)  30.4748050
## yr            6.4538109
## holiday       0.7020538
## workingday  -10.7848799
## temp         51.1712626
## atemp        74.1918968
## hum          -7.5546475
## windspeed    -7.1012553
```

```r
colnames(rbetas)    <- colnames(X)
plotz <- list()
for( j in 1:ncol(X)) {
    Tdat <- rbetas[,j]
qu <- quantile(Tdat, probs = c(0.5, 0.025, 0.975))[2:3]
d <- data.frame( density(Tdat)[[1]], round(density(Tdat)[[2]],6) )
```

```
colnames(d) = c("x" , "y" ); d$area <- d[,1] > qu[1] & d[,1] < qu[2]
J <- ifelse( j == 2 , 9,j ); COLR <- brewer.pal(9,"Set3")[J]

 plotz[[j]] <- ggplot( data = d , aes(x=x , y=y)    ) +
        geom_line( col=COLR  ) + theme_minimal()+
    geom_ribbon(data = d[which(d$area == T),], aes(x, ymin=0 , ymax=y ),
         fill=COLR , alpha = .15) + ylab("") +
    theme(legend.position="none", axis.title.x=element_blank(),
        axis.title.y=element_blank(),
         plot.title = element_text(hjust = 0.5)   )+
    ggtitle( colnames(rbetas)[j] )
 }


multiplot(plotz[[1]], plotz[[2]], plotz[[3]], plotz[[4]],plotz[[5]], plotz[[6]], plotz[[7]], plotz[[8]], co
```

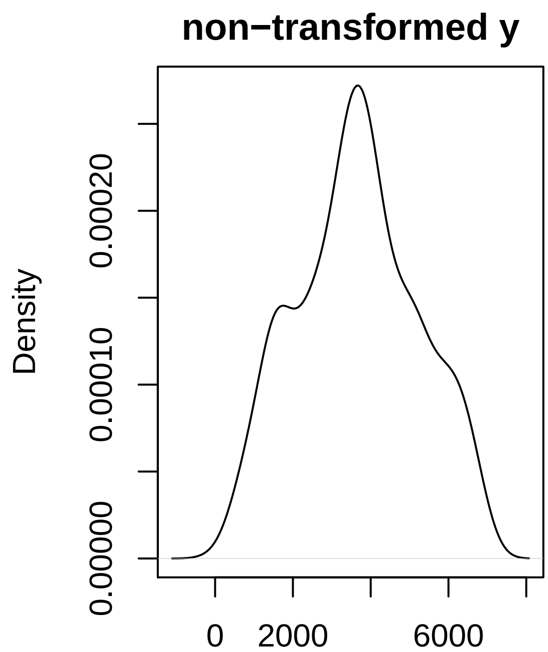

(Intercept)
0.2
0.1
0.0
20
24
28
temp
0.04
0.03
0.02
0.01
0.00
−20
0
20
40
yr
0.75
0.50
0.25
0.00
3
4
5
6
atemp
0.03
0.02
0.01
0.00
0
20
40
60
80
holiday
0.3
0.2
0.1
0.0
−10.0
−7.5
−5.0
−2.5
0.0
hum
0.25
0.20
0.15
0.10
0.05
0.00
−20
−15
−10
workingday
0.8
0.6
0.4
0.2
0.0
−14
−13
−12
−11
windspeed
0.10
0.05
0.00
−30
−20
−10

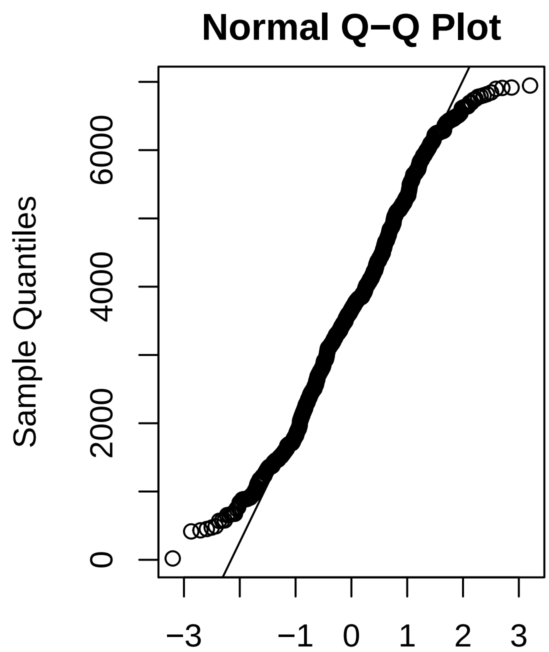
```
## Y
for( i in 1:Nsim) {Yc[  i,] <- rmvnorm(1,  (X %*%rbetas[i,])^2 , sigma = rsig[1] * diag(730) )}
```

```
## Y
ycplots <- list()
ycsums <- list()

DATES <-  seq( as.Date("2011-01-01"), as.Date("2012-12-31"), by="+1 day")
DATES <-  DATES[-which( DATES  == "2012-02-29") ]
rownames(Yc) <- gsub(" ","",as.character(as.yearqtr(DATES)))

for( i in 1:length( unique(rownames(Yc)) ) ){

    yq <- unique(rownames(Yc))[i]
    yq <-  Yc[ which( rownames(Yc) == yq ) , ]
    yq <-  colSums(yq) / dim(yq)[1]
    ycsums[[i]] <- summary(yq)

    qu <- quantile(yq, probs = c(0.5, 0.025, 0.975))[2:3]
    d <- data.frame( density(yq)[[1]], round(density(yq)[[2]],6) )
    colnames(d) = c("x" , "y" ); d$area <- d[,1] > qu[1] & d[,1] < qu[2]
    J <- ifelse( i == 6 , 9,i ); COLR <- brewer.pal(9,"Set3")[J]
    COLR <- brewer.pal(9,"Set1")[J]


 ycplots[[i]] <- ggplot( data = d , aes(x=x , y=y)    ) +
        geom_line( col=COLR  ) + theme_minimal()+
    geom_ribbon(data = d[which(d$area == T),], aes(x, ymin=0 , ymax=y ),
         fill=COLR , alpha = .15) + ylab("") +
    theme(legend.position="none", axis.title.x=element_blank(),
        axis.title.y=element_blank(),
         plot.title = element_text(hjust = 0.5)   )+
    ggtitle(  unique(rownames(Yc))[i] )
 }

ycsums <- as.matrix(do.call(rbind, ycsums))
ycsums <- as.data.frame(t(ycsums))
colnames(ycsums) <- paste0("Yc",unique(rownames(Yc)))
```

We will conduct a similar analysis on registered users y_r.

```
par(mfrow=c(1,2) ,  mar=c(2.1,4.1,2.1,2.1) )
ty <-  (  cb[,2] ); plot(density(ty), main="non-transformed y"); qqnorm(ty); qqline(ty )
```






```
n <- nrow(cb )
Nsim <- 20000
y1 <- ( cb[,2] )
X <- as.matrix( cbind(1, cb[,-(1:2)] ) )
colnames(X) <- c("(Intercept)",colnames(cb[,-(1:2)] ))
p <- ncol(X)

bhat <-  (solve(t(X)%*%X)%*%(t(X)%*%y1))
#fit <- ( lm( y1~X[,-1]  )); coef(fit)
SSY <- SSYr <- t(y1 - X%*%bhat)%*%(y1 - X%*%bhat); #anova(fit)
#vcov( fit )
rbetas <- matrix(0, nrow = Nsim, ncol = p)
XtXi    <- solve(t(X)%*%X)

set.seed(235)
rsig    <- rinvgamma(Nsim , (n-p)/2, (1/2)*SSY)
for(i in 1:Nsim){
    CovX    <- rsig[i]*XtXi
    rbetas[i,] <- c(rmvnorm(1, mean = bhat, sigma = CovX ) )
}

yrrbetas <- rbetas
rbMat   <- apply(rbetas, 2, quantile, probs = c(0.5, 0.025, 0.975))
colnames(rbMat) <- colnames(X) #coef(fit)
rbMat <- as.data.frame( round(t(rbMat), 4) )
rbMat$zero <- rbMat[,2] < 0 & rbMat[,3] > 0
rbMat
```

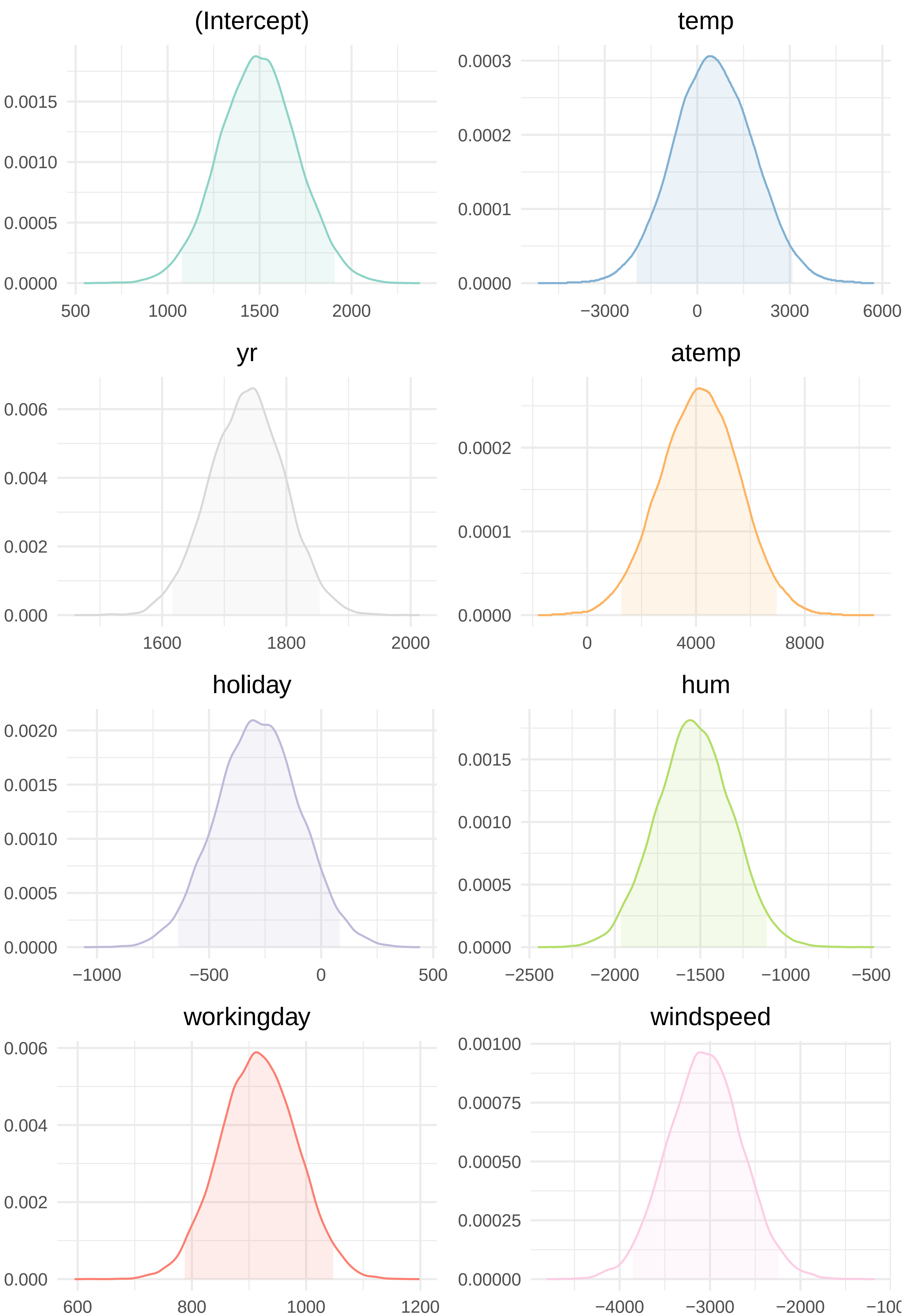
```
##                    50%       2.5%      97.5%  zero
## (Intercept)  1491.7597  1076.0464  1910.4538 FALSE
## yr           1737.2919  1616.2641  1853.9756 FALSE
## holiday      -280.0083  -641.4487    85.9197  TRUE
## workingday    916.6031   786.9039  1047.6404 FALSE
## temp          514.8231 -1977.2392  3092.5738  TRUE
## atemp        4145.0115  1248.9099  6986.4755 FALSE
## hum         -1537.9798 -1965.6322 -1109.7543 FALSE
```

```
## windspeed   -3055.3567 -3855.7252 -2241.6183 FALSE
```

```r
rbsum <- ( apply( rbetas , 2 , summary) )
colnames(rbsum )    <- colnames(X)
 rbsum <- t( rbsum ); rbsum
```

```
##                  Min.    1st Qu.     Median       Mean    3rd Qu.
## (Intercept)   623.6666  1348.7517  1491.7597  1492.2772  1633.5407
## yr           1481.6980  1695.3215  1737.2919  1736.5189  1777.6653
## holiday      -989.2571  -407.1002  -280.0083  -279.7465  -154.2292
## workingday    619.5057   871.3246   916.6031   916.7404   962.5996
## temp        -4686.3344  -353.4328   514.8231   533.0449  1413.6523
## atemp       -1267.1386  3147.5691  4145.0115  4134.6134  5130.2993
## hum         -2368.6096 -1685.9014 -1537.9798 -1536.9066 -1388.3511
## windspeed   -4649.4720 -3333.4716 -3055.3567 -3055.9625 -2781.6229
##                  Max.
## (Intercept)  2292.8458
## yr           1991.7146
## holiday       371.7253
## workingday   1173.7215
## temp         5245.8594
## atemp        9998.8268
## hum          -565.5147
## windspeed   -1333.5393
```

```r
colnames(rbetas)    <- colnames(X)
plotzz <- list()
for( j in 1:ncol(X)) {
    Tdat <- rbetas[,j]
qu <- quantile(Tdat, probs = c(0.5, 0.025, 0.975))[2:3]
d <- data.frame( density(Tdat)[[1]], round(density(Tdat)[[2]],6) )
colnames(d) = c("x" , "y" ); d$area <- d[,1] > qu[1] & d[,1] < qu[2]
J <- ifelse( j == 2 , 9,j ); COLR <- brewer.pal(9,"Set3")[J]

 plotzz[[j]] <- ggplot( data = d , aes(x=x , y=y)    ) +
        geom_line( col=COLR  ) + theme_minimal()+
    geom_ribbon(data = d[which(d$area == T),], aes(x, ymin=0 , ymax=y ),
         fill=COLR , alpha = .15) + ylab("") +
    theme(legend.position="none", axis.title.x=element_blank(),
        axis.title.y=element_blank(),
         plot.title = element_text(hjust = 0.5)   )+
    ggtitle( colnames(rbetas)[j] )
 }


multiplot(plotzz[[1]], plotzz[[2]], plotzz[[3]], plotzz[[4]],plotzz[[5]], plotzz[[6]], plotzz[[7]], plotzz[
```


(Intercept)
0.0015
0.0010
0.0005
0.0000
500
1000
1500
2000
temp
0.0003
0.0002
0.0001
0.0000
−3000
0
3000
6000
yr
0.006
0.004
0.002
0.000
1600
1800
2000
atemp
0.0002
0.0001
0.0000
0
4000
8000
holiday
0.0020
0.0015
0.0010
0.0005
0.0000
−1000
−500
0
500
hum
0.0015
0.0010
0.0005
0.0000
−2500
−2000
−1500
−1000
−500
workingday
0.006
0.004
0.002
0.000
600
800
1000
1200
windspeed
0.00100
0.00075
0.00050
0.00025
0.00000
−4000
−3000
−2000
−10

```r
## Y
for( i in 1:Nsim) {Yr[  i,] <- rmvnorm(1,  (X %*%rbetas[i,])^2 , sigma = rsig[1] * diag(730) )}
```

```r
yrplots <- list()
yrsums <- list()

DATES <-  seq( as.Date("2011-01-01"), as.Date("2012-12-31"), by="+1 day")
DATES <-  DATES[-which( DATES  == "2012-02-29") ]
rownames(Yr) <- gsub(" ","",as.character(as.yearqtr(DATES)))

for( i in 1:length( unique(rownames(Yr)) ) ){

    yq <- unique(rownames(Yr))[i]
    yq <-  Yr[ which( rownames(Yr) == yq ) , ]
    yq <-  colSums(yq) / dim(yq)[1]
    yrsums[[i]] <- summary(yq)

    qu <- quantile(yq, probs = c(0.5, 0.025, 0.975))[2:3]
    d <- data.frame( density(yq)[[1]], round(density(yq)[[2]],6) )
    colnames(d) = c("x" , "y" ); d$area <- d[,1] > qu[1] & d[,1] < qu[2]
    J <- ifelse( i == 6 , 9,i ); COLR <- brewer.pal(9,"Set3")[J]
    COLR <- brewer.pal(9,"Set1")[J]


 yrplots[[i]] <- ggplot( data = d , aes(x=x , y=y)    ) +
        geom_line( col=COLR  ) + theme_minimal()+
    geom_ribbon(data = d[which(d$area == T),], aes(x, ymin=0 , ymax=y ),
         fill=COLR , alpha = .15) + ylab("") +
    theme(legend.position="none", axis.title.x=element_blank(),
        axis.title.y=element_blank(),
         plot.title = element_text(hjust = 0.5)   )+
    ggtitle(  unique(rownames(Yc))[i] )
 }


yrsums <- as.matrix(do.call(rbind, yrsums))
yrsums <- as.data.frame(t(yrsums))
colnames(yrsums) <- paste0("Yc",unique(rownames(Yr)))
```

We can now compare the two outcome variables for the model. For simplicity, I have aggregated the in-model predictions. We know have their means by quarter. In other words, $X\hat{\beta}$ is an 730×20000 matrix. I have condensed the days into quarters by taking the mean count of users for each quarter. So we have the distribution of the mean (20000 observations) for the eight quarters in 2011-2012. It should be noted that the user count for casual riders is $\left(X\hat{\beta}\right)^2$ because of the square root transformation on y_c.

The casual user's predictions will always come first. Here are the summaries:

```r
ycsums
```

```
##          Yc2011Q1 Yc2011Q2  Yc2011Q3 Yc2011Q4 Yc2012Q1 Yc2012Q2 Yc2012Q3
## Min.     240.7795 730.8946  988.2821 408.9047 520.9263 1070.155 1333.846
## 1st Qu. 275.0686 791.7344 1069.3732 453.3511 569.2863 1154.661 1439.620
## Median  283.2841 804.3852 1087.1334 462.8101 580.2906 1170.137 1459.097
## Mean    283.5188 804.4816 1087.2345 462.9331 580.5261 1170.141 1459.203
## 3rd Qu. 291.8607 817.2792 1105.0561 472.3860 591.6347 1185.728 1479.177
```

```
## Max.    337.3983 882.0401 1183.1397 516.3604 648.8412 1271.952 1580.308
##         Yc2012Q4
## Min.    597.9138
## 1st Qu. 647.4756
## Median  659.0046
## Mean    659.1183
## 3rd Qu. 670.5863
## Max.    729.4159
```

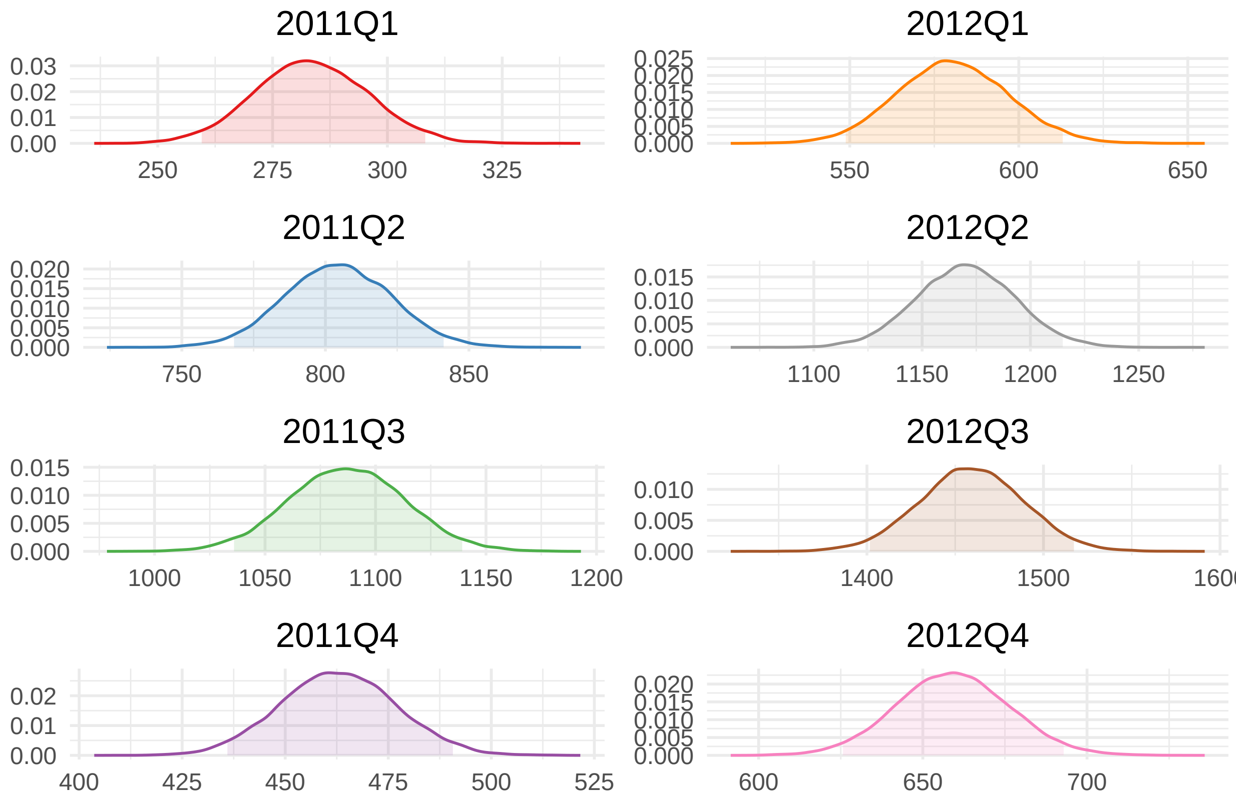
```
yrsums
```

```
##         Yc2011Q1 Yc2011Q2 Yc2011Q3 Yc2011Q4 Yc2012Q1 Yc2012Q2 Yc2012Q3
## Min.    1627.320 2887.764 3392.512 2181.457 3672.147 4804.668 5214.106
## 1st Qu. 1797.480 3051.660 3577.752 2330.840 3861.369 4937.742 5378.398
## Median  1835.953 3082.888 3614.206 2363.249 3895.246 4968.084 5414.482
## Mean    1835.747 3082.986 3614.591 2362.913 3895.241 4968.317 5414.259
## 3rd Qu. 1874.349 3113.630 3651.590 2395.027 3929.082 4998.924 5450.454
## Max.    2046.666 3264.729 3816.724 2544.246 4091.477 5156.733 5636.349
##         Yc2012Q4
## Min.    3871.961
## 1st Qu. 4034.964
## Median  4067.404
## Mean    4067.332
## 3rd Qu. 4099.716
## Max.    4255.430
```

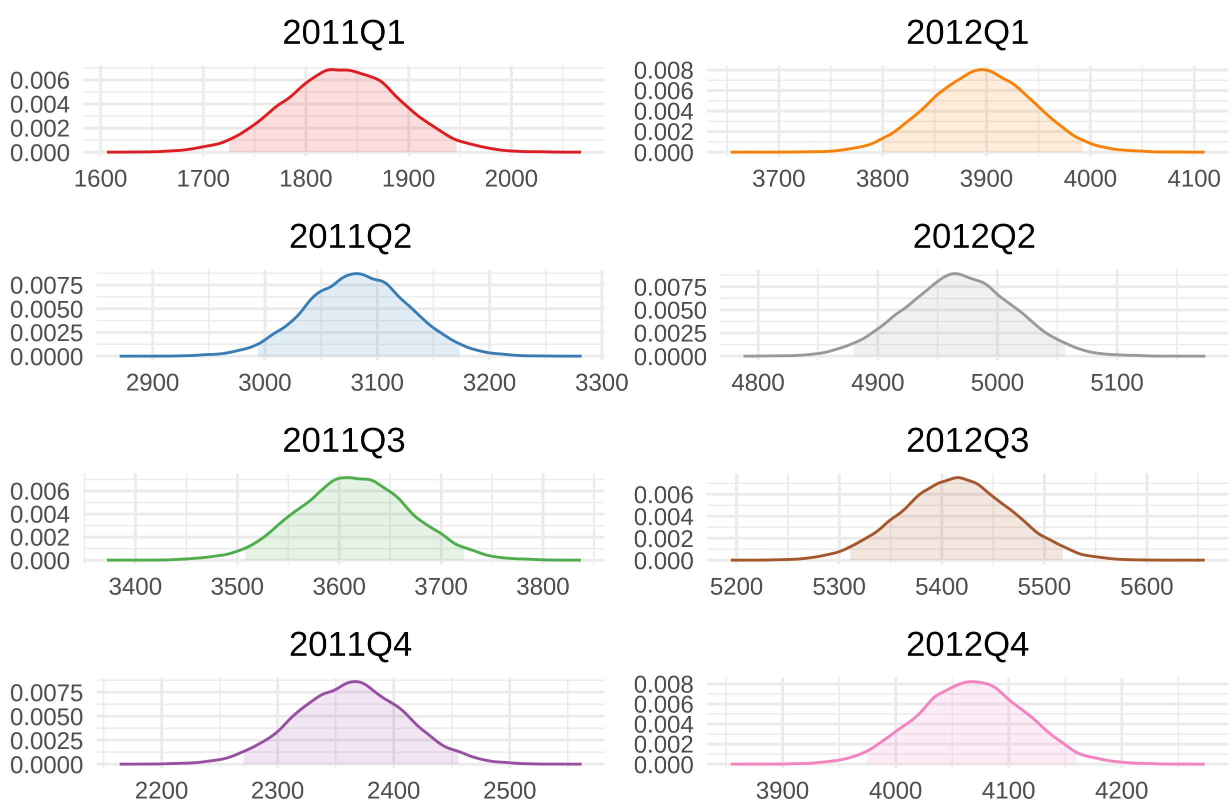
There are many more registered riders than there are casual riders. They have similar jumps and falls in ridership as it gets warmer/colder. It is interesting to note that the jump for casual riders in Q1 to Q3 is about triple or quadruple. The jump for registered users in the same time period is less than double.

Also it is clear that Capital Bikeshare increased in popularity from 2011 to 2012. It appears ridership almost doubles.

```
multiplot(ycplots[[1]], ycplots[[2]], ycplots[[3]], ycplots[[4]],ycplots[[5]], ycplots[[6]], ycplots[[7]],
```




```
multiplot(yrplots[[1]], yrplots[[2]], yrplots[[3]], yrplots[[4]],yrplots[[5]], yrplots[[6]], yrplots[[7]],
```



Prediction (from Gelman pg. 357):

```r
pred <- c(1,1,0,1,.344348,0.34847,0.804783,0.179117)

ycp <- ycpreds <- apply( ycrbetas , 1 , function(B) (pred %*% B)^2 )
ycpreds <- ycpreds*0

set.seed(2011)
rsig    <- rinvgamma(Nsim , (n-p)/2, (1/2)*SSYc)
for(i in 1:Nsim){
    ycpreds[i] <- (rnorm(1, mean = ycp[i], sqrt(rsig[i])  ) )
}

summary(ycpreds);quantile(ycpreds, probs = c(0.5, 0.025, 0.975))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   207.2   277.3   290.0   290.1   302.4   371.2

##      50%     2.5%    97.5%
## 290.0186 254.3797 327.0514
```

```r
yrp <-  yrpreds <- apply( yrrbetas , 1 , function(B) pred %*% B )
yrpreds <- yrpreds*0

set.seed(2012)
rsig    <- rinvgamma(Nsim , (n-p)/2, (1/2)*SSYr)
for(i in 1:Nsim){
    yrpreds[i] <- c(rnorm(1, mean = yrp[i], sqrt(rsig[i])  ) )
}

summary(yrpreds); quantile(yrpreds, probs = c(0.5, 0.025, 0.975))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   272.8  3431.1  3991.2  3988.3  4536.0  7914.9

##      50%     2.5%    97.5%
## 3991.219 2402.704 5582.055
```