Daanish Ahsan
Advanced Topics in Statistics
Candidate Number: 124070

Assignment 2

1)

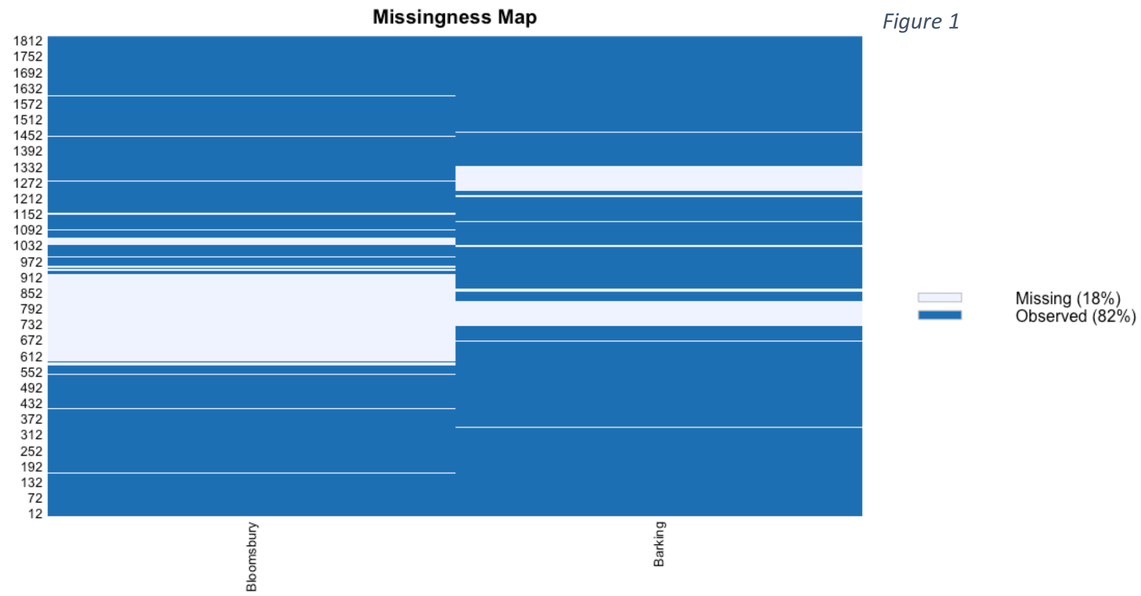


Figure 1

In Bloomsbury the average level of of PM10 pollution is 5.5, with a maximum of 77.7 and an average level of 22.09; there are 427 NA values in the Bloomsbury dataset. In terms of NA values by year in Bloomsbury, there are 12 in 2000 and 2001, 240 in 2002, 154 in 2003, and 9 in 2004. In Barking the average level of PM10 pollution is 21.53, with a minimum of 3.8, maximum of 71.8, and 232 NA values. By year, Barking has 3 NA values in 2000, 104 in 2001, 113 in 2002, 11 in 2003, and 1 in 2004. Bloomsbury has more missing values, with the NA values being very low for 2 years, rising suddenly in the next 2, before falling very low again in 2004. Barking starts very low also, and then rises quick for the next two years, before pollution falls very low for the last 2 years. Clearly, for both areas, the patterns of missing data change over time. This is shown by the missingess map in Figure 1, by the significant chunk of 2002 missing for Bloomsbury, and smaller chunks missing for Barking in 2001 and 2002.

Figure 2

2)

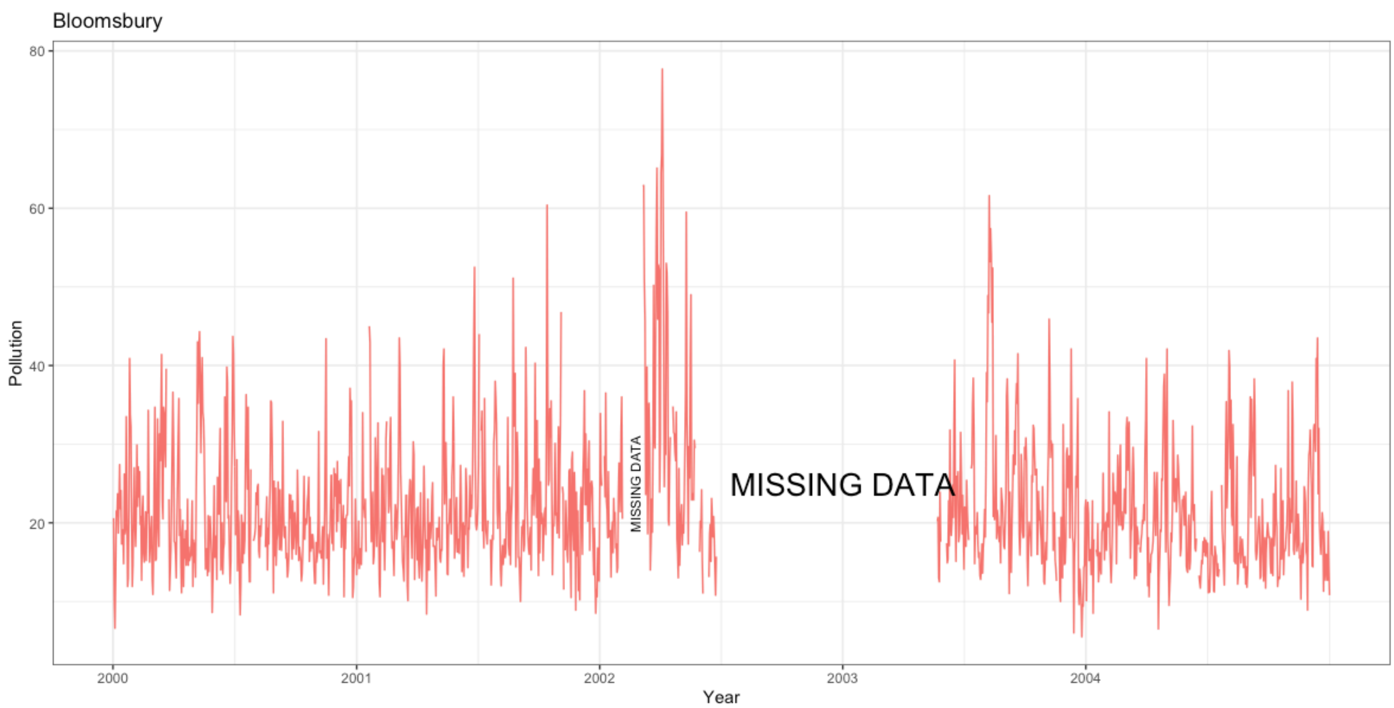
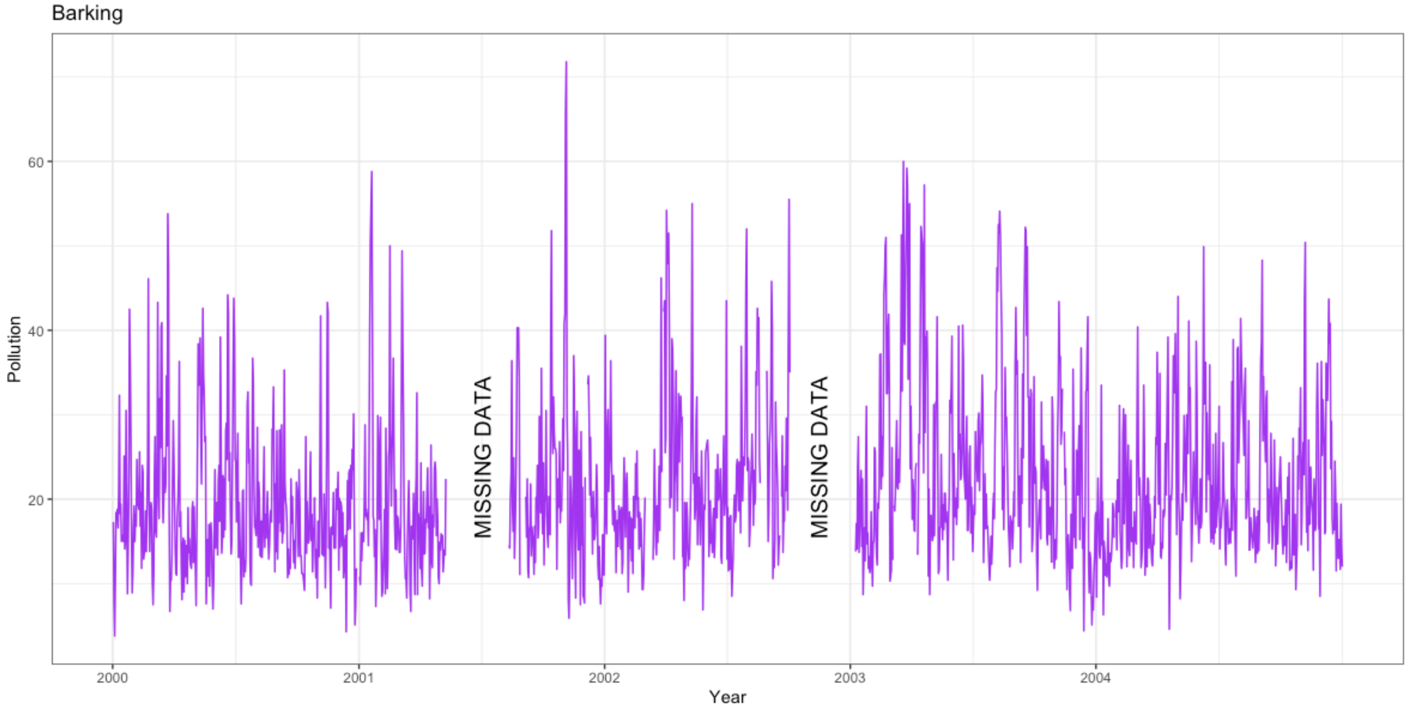


Figure 3



3)

Figure 4

Figure 4 shows the locations of Barking and Bloomsbury on a map of London; Barking is red and Bloomsbury is blue. Bloomsbury is located in the West End, a very tourist heavy area, whilst Barking is in East London. Bloombury's higher average, minimum, and maximum pollution may be explained by the fact that it is a very tourist heavy area, and also has multiple universities and museums located there.

Daanish Ahsan
Advanced Topics in Statistics
Candidate Number: 124070
4/5/6)

Figure 5

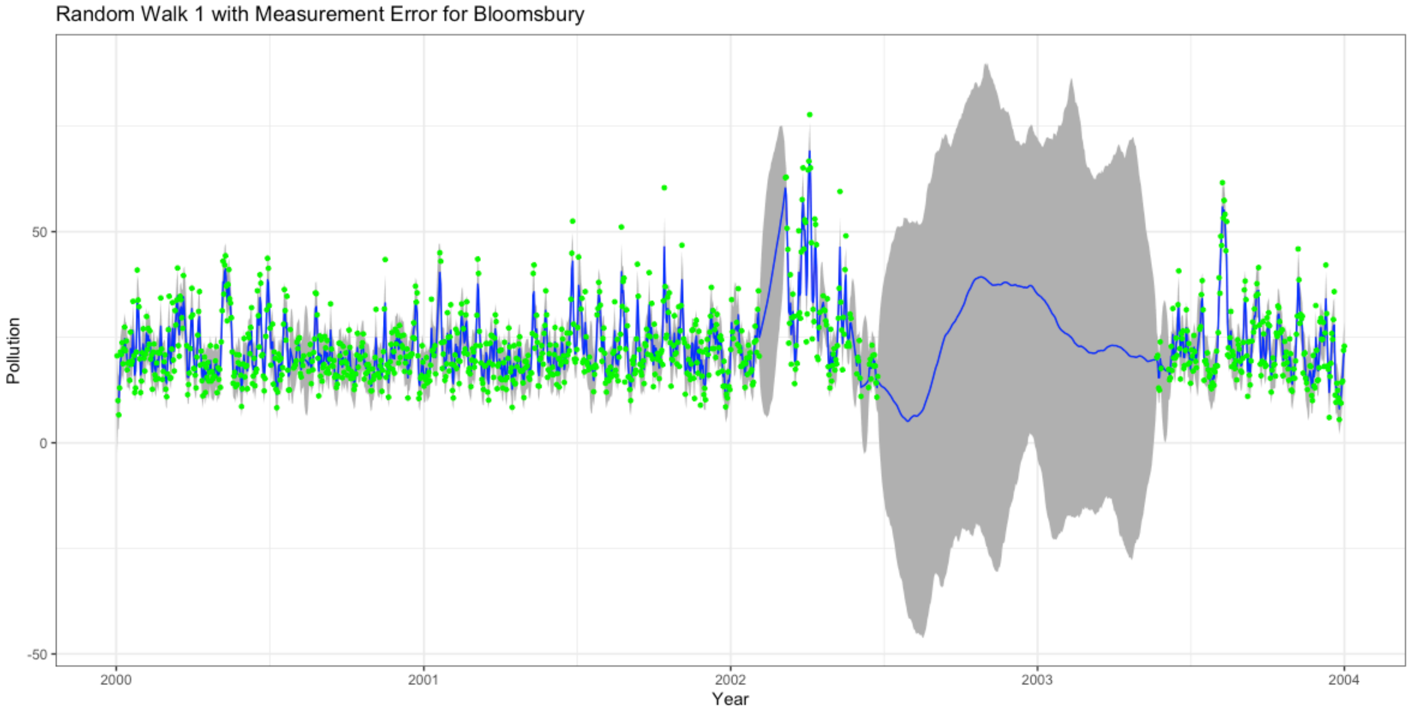


Figure 6

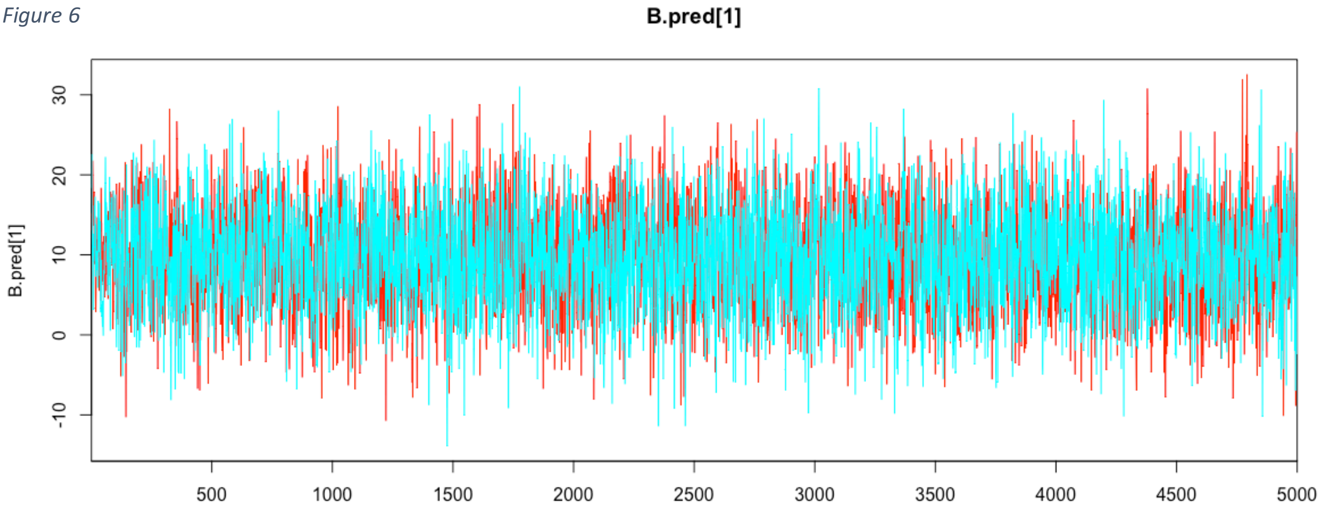
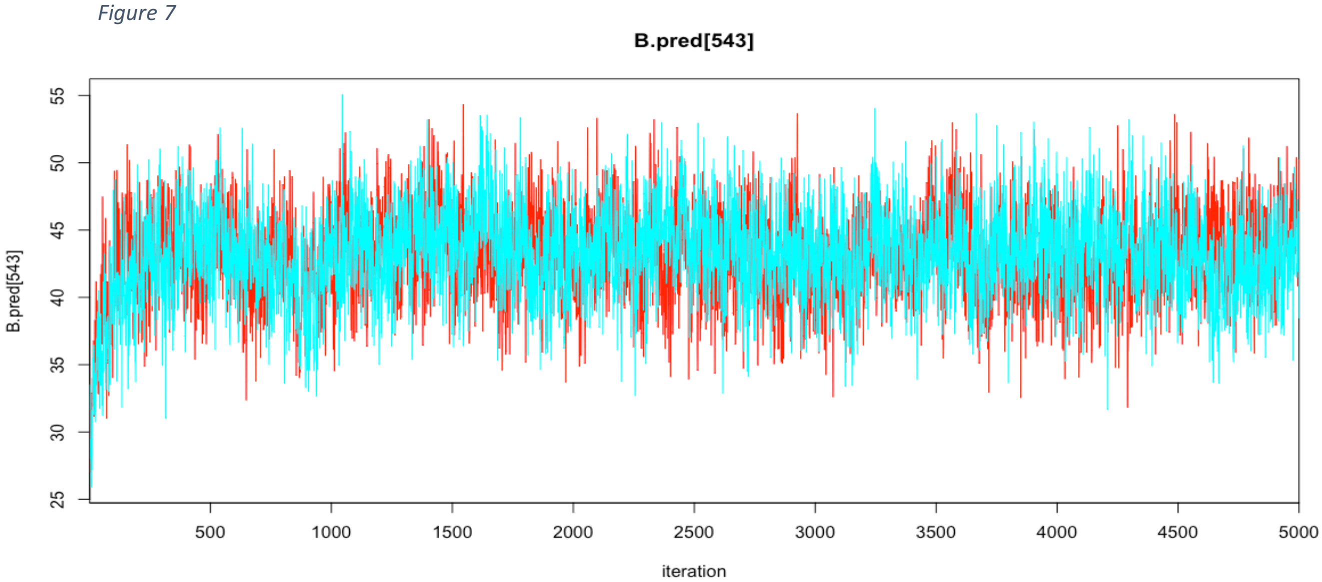


Figure 7



Daanish Ahsan
Advanced Topics in Statistics
Candidate Number: 124070

Gelman Diagnostic
for Bloomsbury
Random Walk 1

	Point est.	Upper C.I.
B.pred[1]	0.9999728	1.0002011
B.pred[10]	1.0006330	1.0006331
B.pred[100]	0.9999436	1.0000567
B.pred[1000]	1.7289491	3.4612309
Bloomsbury[1000]	1.7079327	3.3559620
Bloomsbury[1001]	1.6929037	3.3987269
Bloomsbury[1002]	1.6772799	3.4128634
deviance	1.0051572	1.0085559
MSErw1	1.0046632	1.0071221
sigma.v2	1.0045029	1.0057244
sigma.w2	1.0070508	1.0129532
tau.v	1.0077037	1.0140151
tau.w	1.0071243	1.0071255

In terms of convergence in this model, the traceplots mostly seem to converge. Looking at the Gelman diagnostic is more helpful, and whilst the above only shows the Gelman values for a few parameters, looking through the whole output shows there are many parameters which are less than 1.1, but also many slightly above 1.1, which overall indicates a lack of overall model convergence, despite the converging traceplots. Figure 5 shows the RW process in blue, and the actual data points in green; the model seems to fit the actual data points very well.

Daanish Ahsan
Advanced Topics in Statistics
Candidate Number: 124070

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
B.pred[1]	9.5351559	6.0254504	-2.1767398	5.3859104	9.5414994	13.6447611	21.1738745	1.000955	10000
B.pred[2]	9.7523796	3.4333462	2.9787301	7.4256501	9.7724557	12.0648285	16.4585776	1.001093	7700
B.pred[3]	9.5674463	3.0960575	3.5317300	7.5119024	9.5547174	11.5936914	15.7078534	1.001340	3400
B.pred[4]	13.8085924	3.0230021	7.8204282	11.8084694	13.7699813	15.8101845	19.8553550	1.001084	8100
B.pred[5]	19.0542740	3.0514485	13.0526078	17.0033015	19.0655769	21.0870345	24.9305802	1.001160	5700
MSErw1	4.2156489	0.3718739	3.5837016	3.9836596	4.2002592	4.4180860	4.8953155	1.005011	1600
deviance	5943.3999698	177.7338982	5614.0413316	5834.7302178	5944.2526998	6047.0802465	6260.9211786	1.005442	1600
sigma.v2	17.8142037	3.3864511	12.5964655	15.7178321	17.4851787	19.4524571	24.0729475	1.004820	1800
sigma.w2	26.4197795	4.2013926	18.3849628	24.0150748	26.5760261	29.1190890	34.0025284	1.007962	3800
tau.v	0.0579196	0.0100791	0.0415404	0.0514074	0.0571913	0.0636220	0.0793873	1.004820	1800
tau.w	0.0393427	0.0113523	0.0294096	0.0343417	0.0376279	0.0416405	0.0543923	1.007962	3800

Summary of RW1 model for Bloomsbury

Lack of full model convergence is further seen through the Rhat values in the model summary shown above; in the full output, many Rhat values exceed 1.1, as in the Gelman diagnostic, thus inferring a lack of convergence. In terms of the credible interval widths for the missing data periods, their widths are far larger than for the periods that contain data, which makes sense as the model is predicting between which values the actual data would be. The credible widths also look like they broadly follow the trend of the missing data points.

Daanish Ahsan
Advanced Topics in Statistics
Candidate Number: 124070

7)

Figure 8

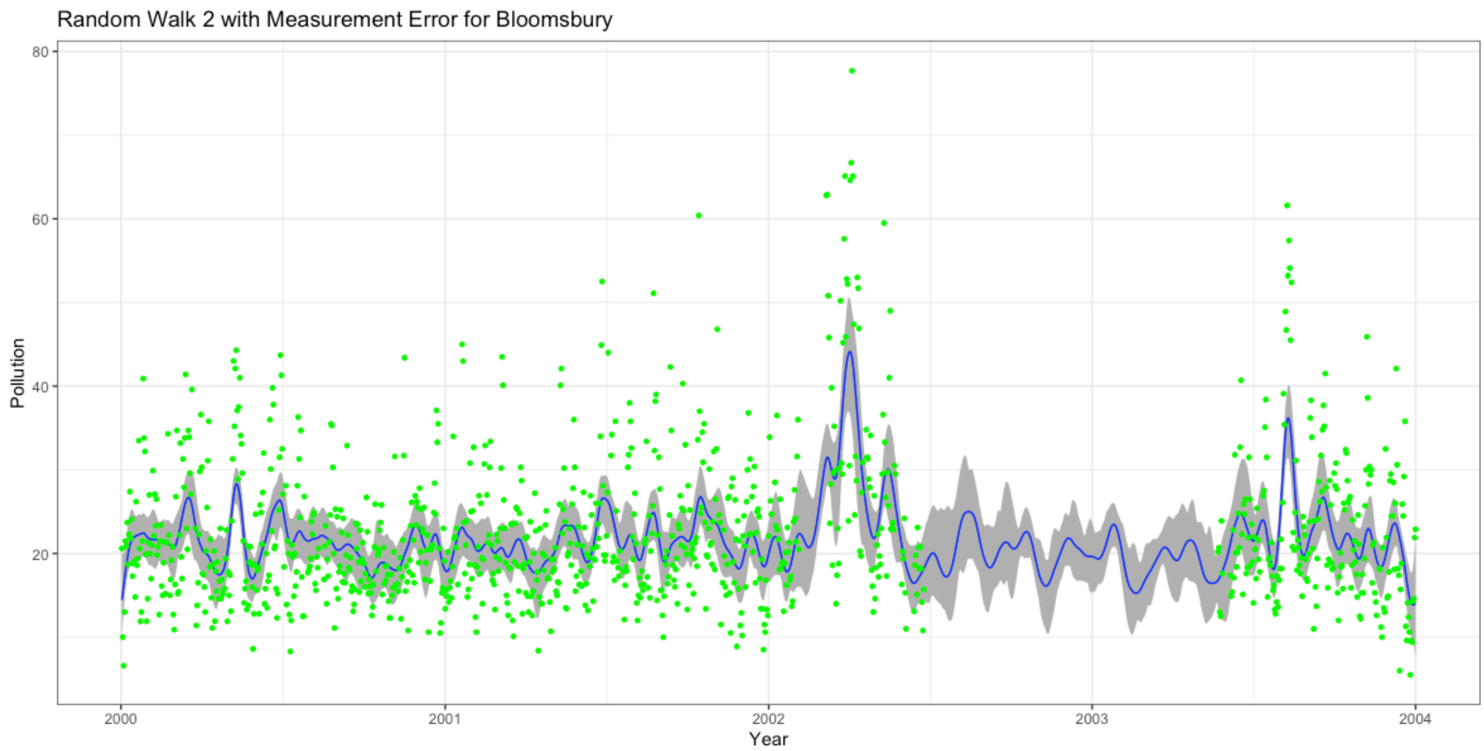


Figure 9

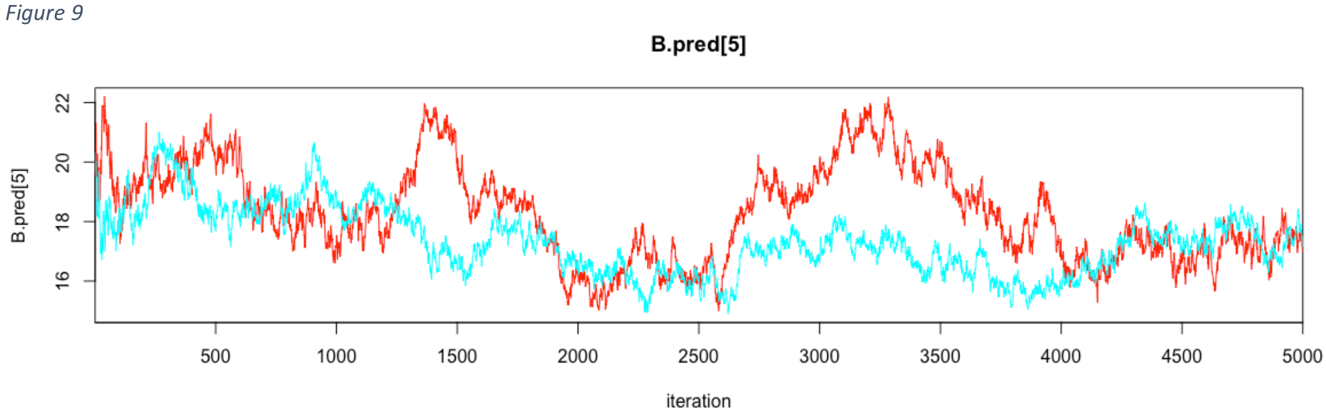
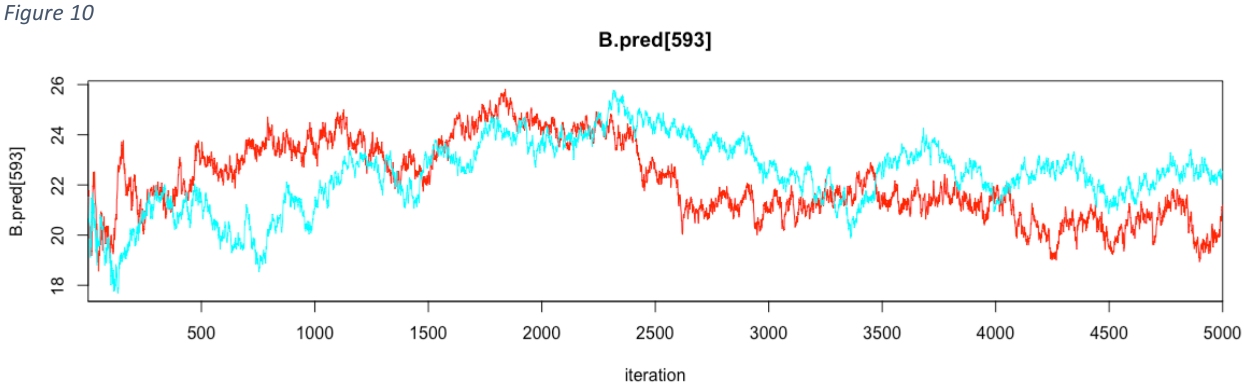


Figure 10



	Point est.	Upper C.I.
B.pred[1]	0.9999728	1.0002011
B.pred[10]	1.0006330	1.0006331
B.pred[100]	0.9999436	1.0000567
B.pred[1000]	1.7289491	3.4612309
Bloomsbury[1000]	1.7079327	3.3559620
Bloomsbury[1001]	1.6929037	3.3987269
Bloomsbury[1002]	1.6772799	3.4128634
deviance	1.0051572	1.0085559
MSErw1	1.0046632	1.0071221
sigma.v2	1.0045029	1.0057244
sigma.w2	1.0070508	1.0129532
tau.v	1.0077037	1.0140151
tau.w	1.0071243	1.0071255

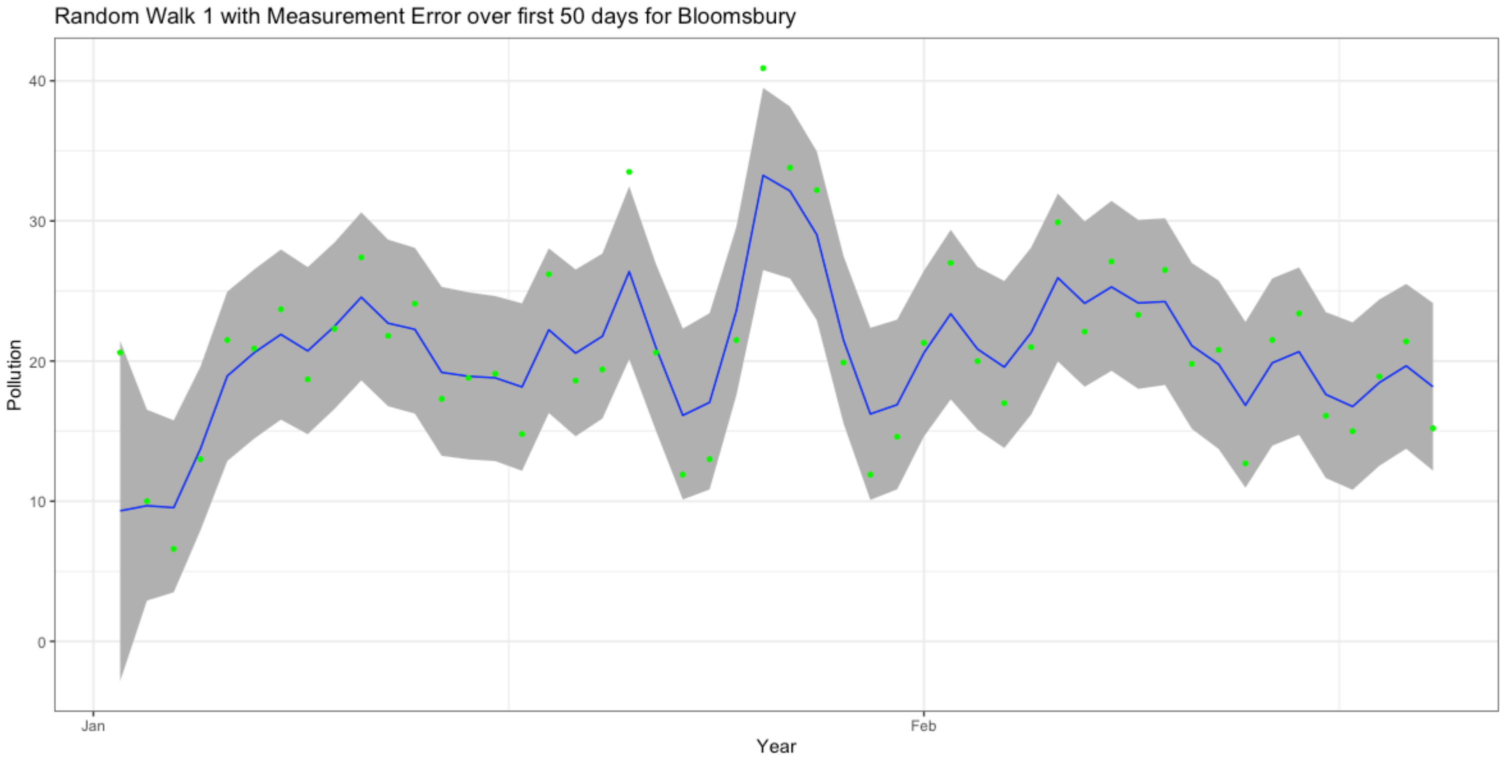
Gelman Diagnostic
for Bloomsbury
Random Walk 2

This model does not seem to fit the data as well as the RW1 model. In terms of convergence, none of the traceplots seem to converge, and the Gelman diagnostic and Rhat values shown in the summary table below, are mostly slightly or significantly greater than 1.1, futher indicating a lack of convergence. So the RW1 model seems a better fit for the data, and converges more than the RW2 model.

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
B.pred[1]	14.4673503	2.7224623	9.4650516	12.6930016	14.2740632	16.1350976	20.3280528	1.005691	10000
B.pred[2]	15.3212318	2.1796168	11.4502964	13.8614416	15.1195607	16.6855362	20.0830207	1.008770	210
B.pred[3]	16.1665134	1.7684078	13.1173384	14.8960099	15.9934570	17.2972689	19.9037572	1.040980	44
B.pred[4]	17.0024204	1.5284351	14.4696791	15.8605422	16.8100306	18.0165180	20.3121579	1.130271	17
B.pred[5]	17.8127860	1.4446502	15.5602723	16.7369574	17.5978077	18.6941486	21.0668192	1.265146	10
MSErw1	7.6061903	0.1516851	7.3233006	7.4990285	7.6045966	7.7082582	7.9147298	1.890807	4
deviance	7180.0695875	33.6852582	7119.3727176	7151.1589032	7183.6085238	7202.4483943	7248.8689546	2.543572	3
sigma.v2	57.8926985	3.1679594	51.9866721	55.6846304	57.7883274	60.0022295	64.3152132	1.406529	7
sigma.w2	0.0953862	0.0813685	0.0503816	0.0616523	0.0813825	0.1059292	0.2169258	1.709127	5
tau.v	0.0173250	0.0009461	0.0155484	0.0166660	0.0173045	0.0179583	0.0192357	1.406529	7
tau.w	12.6361759	4.1979588	4.6098722	9.4402720	12.2876472	16.2199884	19.8485051	1.709127	5

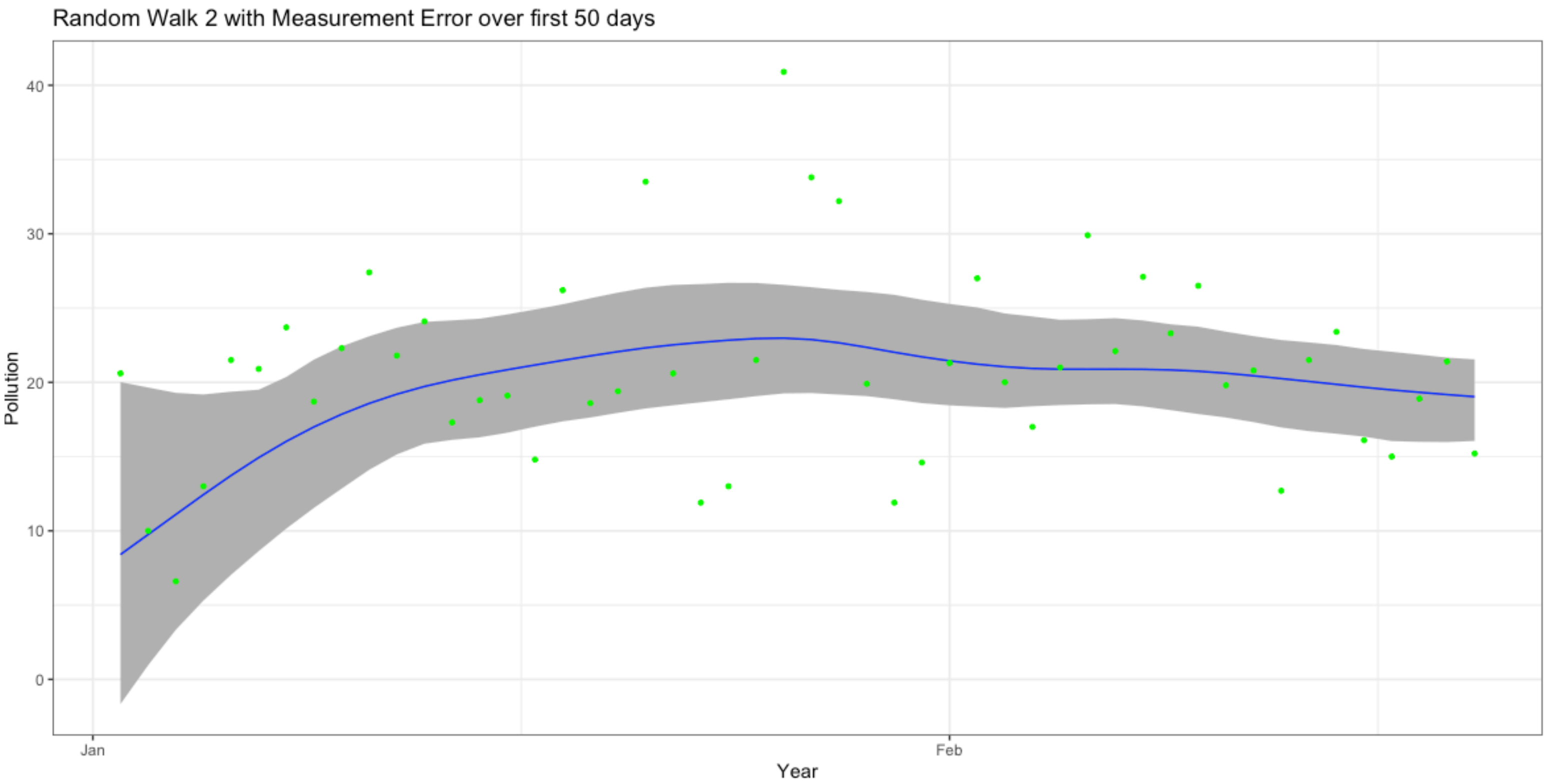
Summary of RW2 model for Bloomsbury

Figure 11



Comparing the first 50 days for each model shows that RW2 has a far greater smoothing effect than RW1

Figure 12



Daanish Ahsan
Advanced Topics in Statistics
Candidate Number: 124070

8/9)

Figure 13

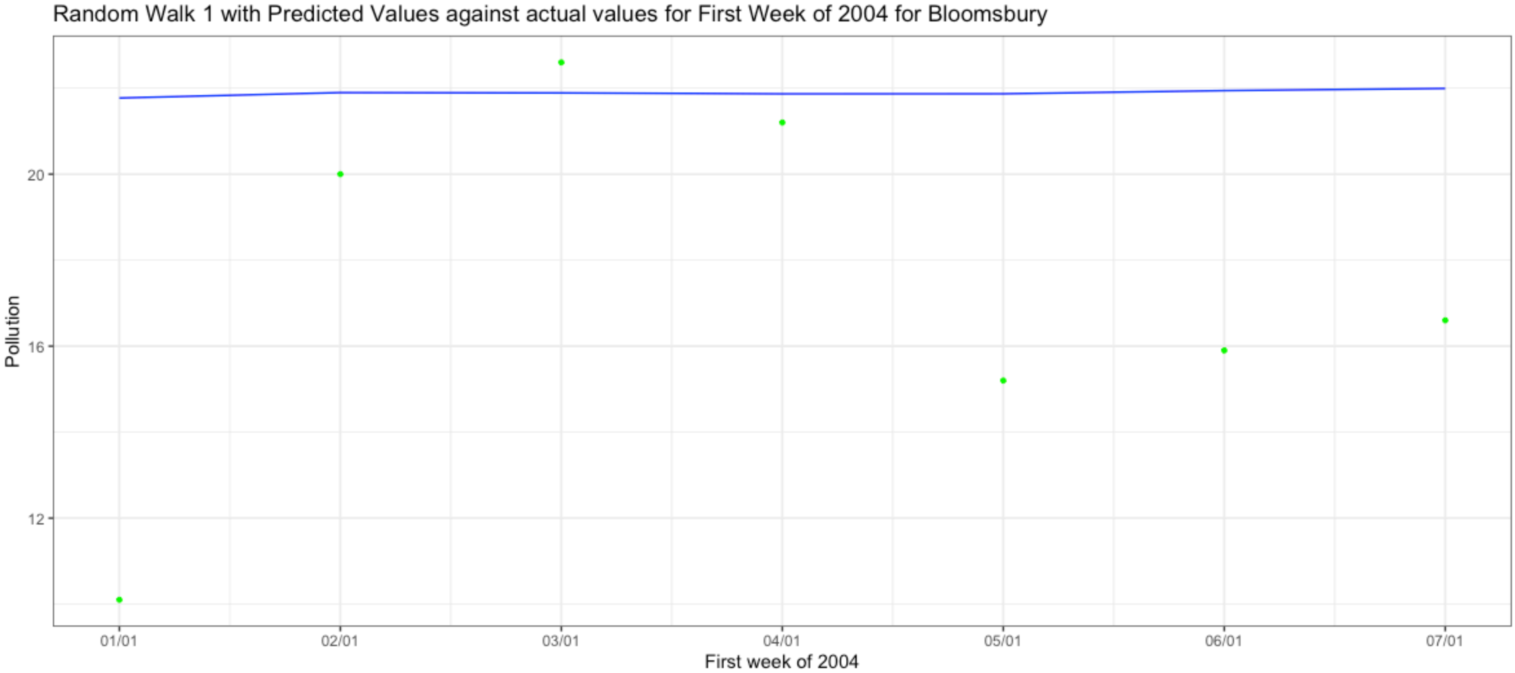
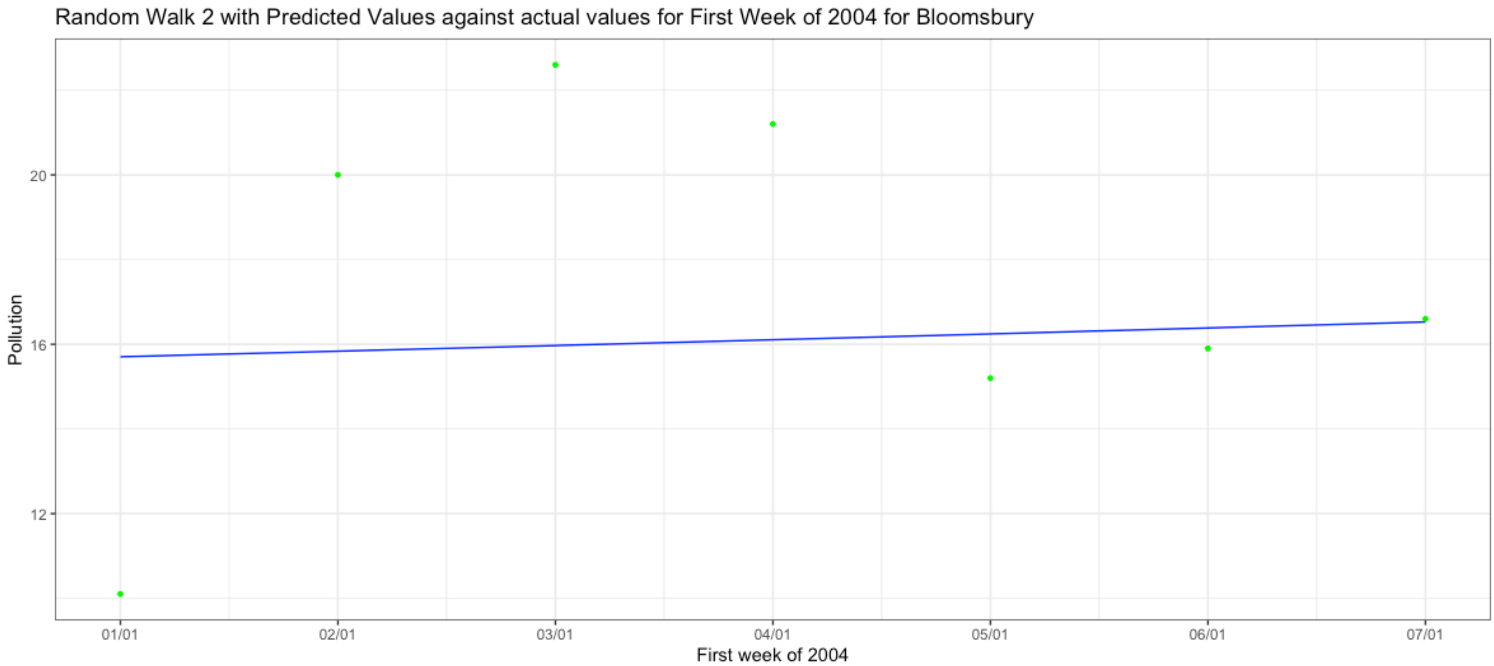


Figure 14



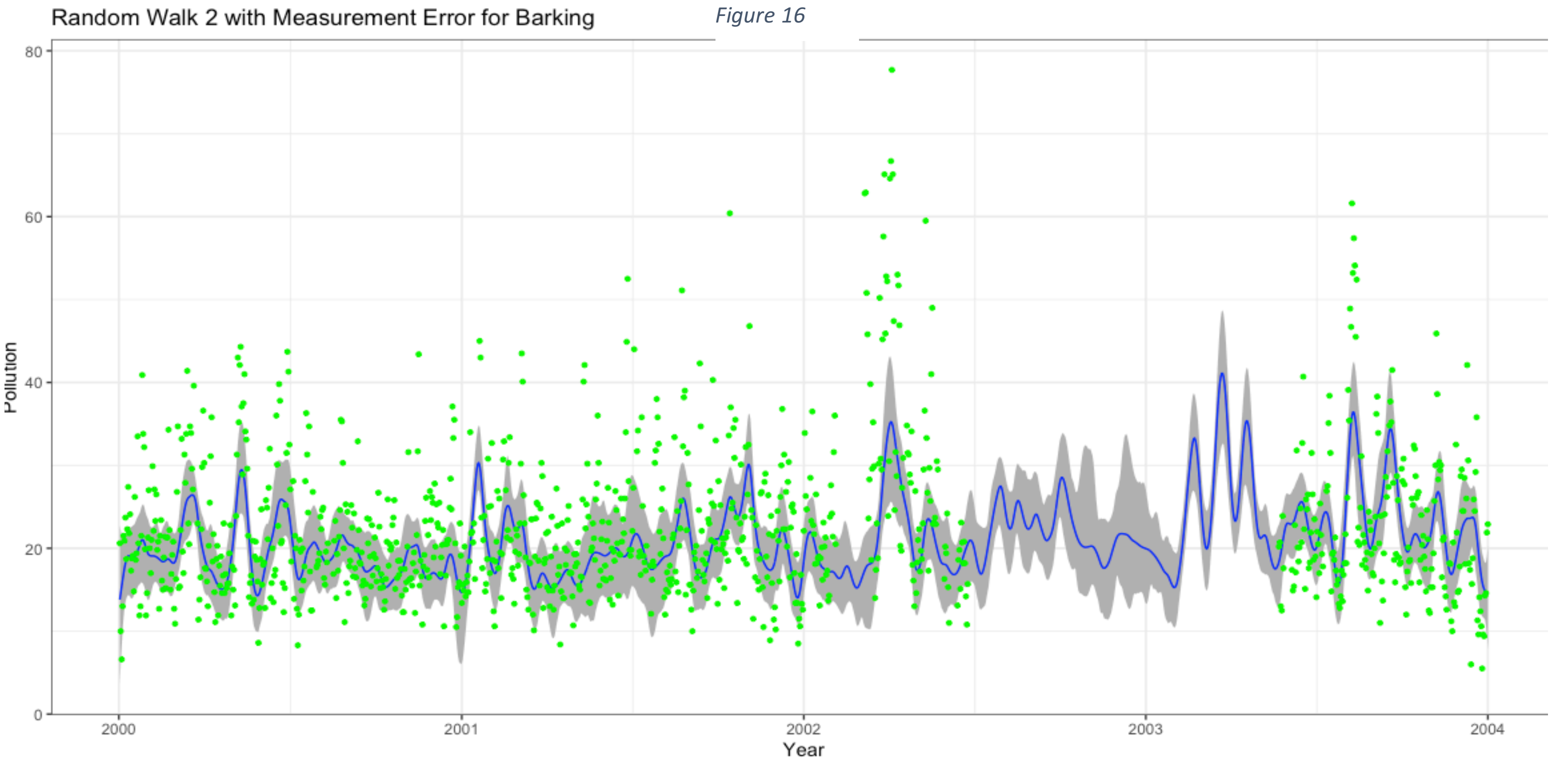
Looking at figures 13 and 14 above, RW2 appears to be closer to 3 of the actual data points than the 3 that RW1 is close to. Yet the mean squared error of RW 1 is 4.25, compared to 7.75 for RW2, suggesting that the RW1 model is better at forecasting. Perhaps more than 7 days are needed to judge if the forecasts are accurate.

Daanish Ahsan
Advanced Topics in Statistics
Candidate Number: 124070

10)

Figure 15

Figure 16



Similarly to the Bloomsbury models, the traceplots show that RW1 for Barking has mostly converged, but for RW2 the traceplots don't converge; these patterns are further confirmed when comparing the Gelman diagnostic values between the two models. In terms of model fit, as for Bloomsbury, RW1 fits the data better than RW2.

Daanish Ahsan
Advanced Topics in Statistics
Candidate Number: 124070

11)

Figure 17

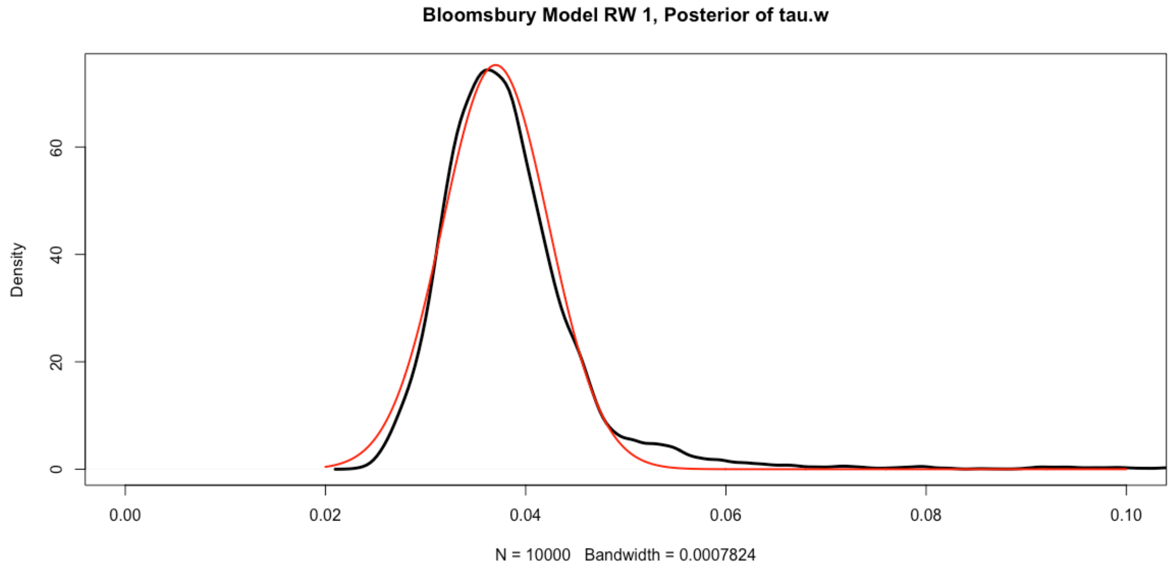


Figure 18

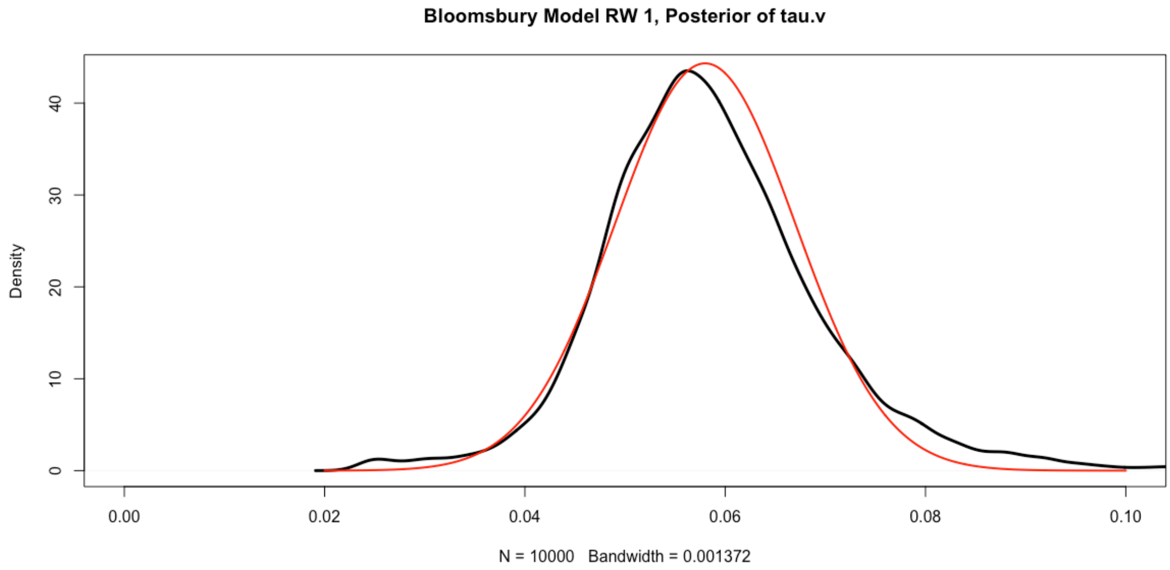
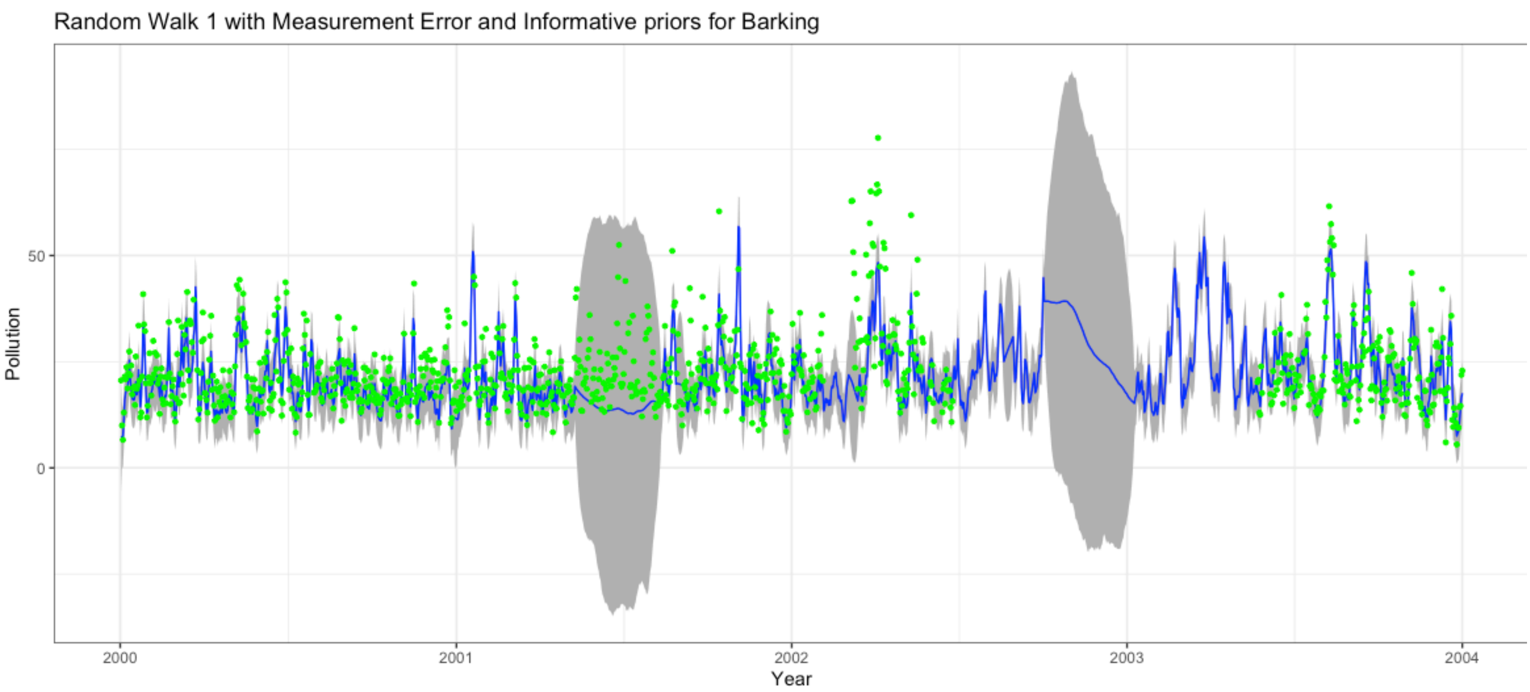


Figure 19



Daanish Ahsan
Advanced Topics in Statistics
Candidate Number: 124070

Figures 17 and 18 show the distributions of tau w and tau v for RW 1, and the red line is adjusting the mean and standard deviations to fit these distributions as closely as possible. Using these distributions and thus informative priors to re run Barking for RW1, is shown in Figure 19. Compared to RW1 with non-informative priors in Figure 15, they both seem to fit the data equally well, but differ more in the missing data trend line, although only slightly, but this also means the credible intervals are slightly different between the two models. In terms of convergence, the RW1 informative priors model's traceplots almost all converge, as they did for the non informative priors RW1 model for Barking. Again, this convergence pattern is indicated by the gelman diagnostic and rhat values in the model summary, which shows that most values are less than 1.1.

Figure 20

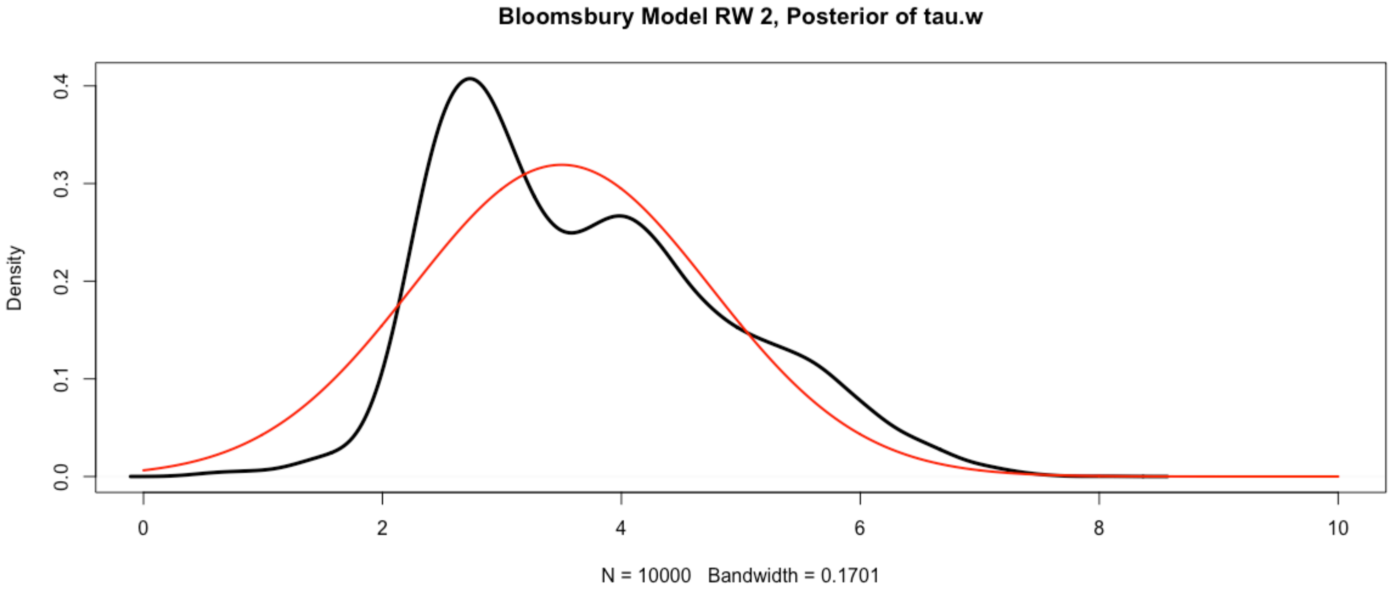


Figure 21

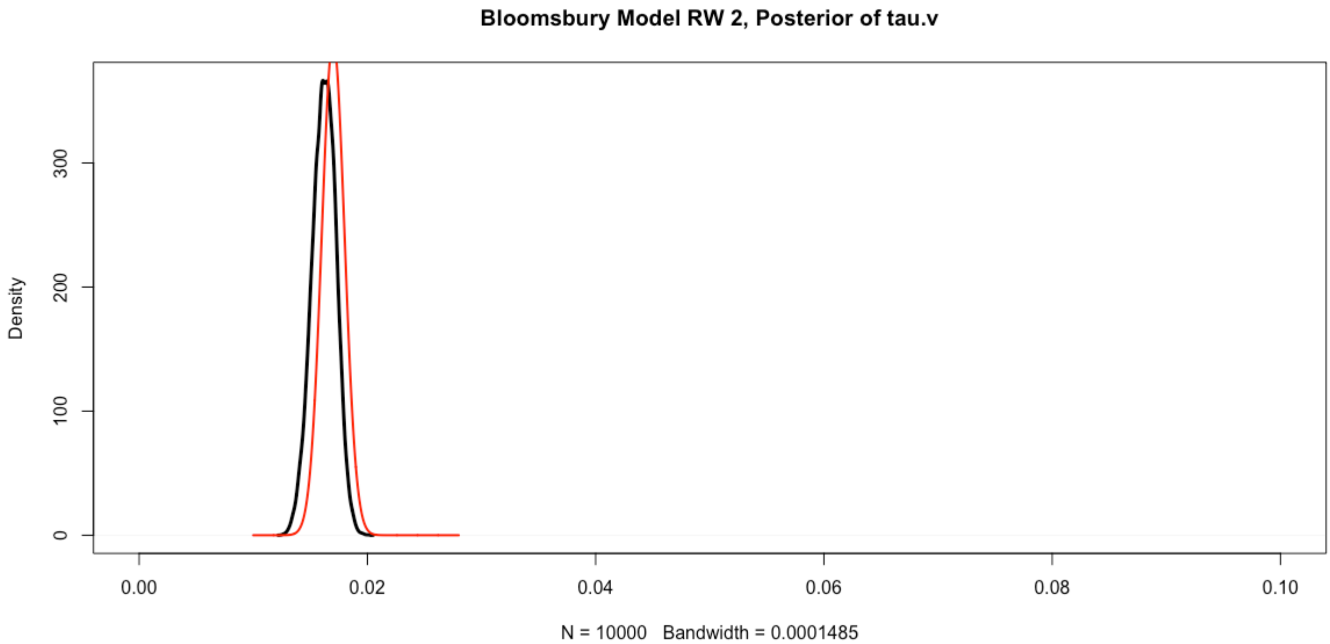
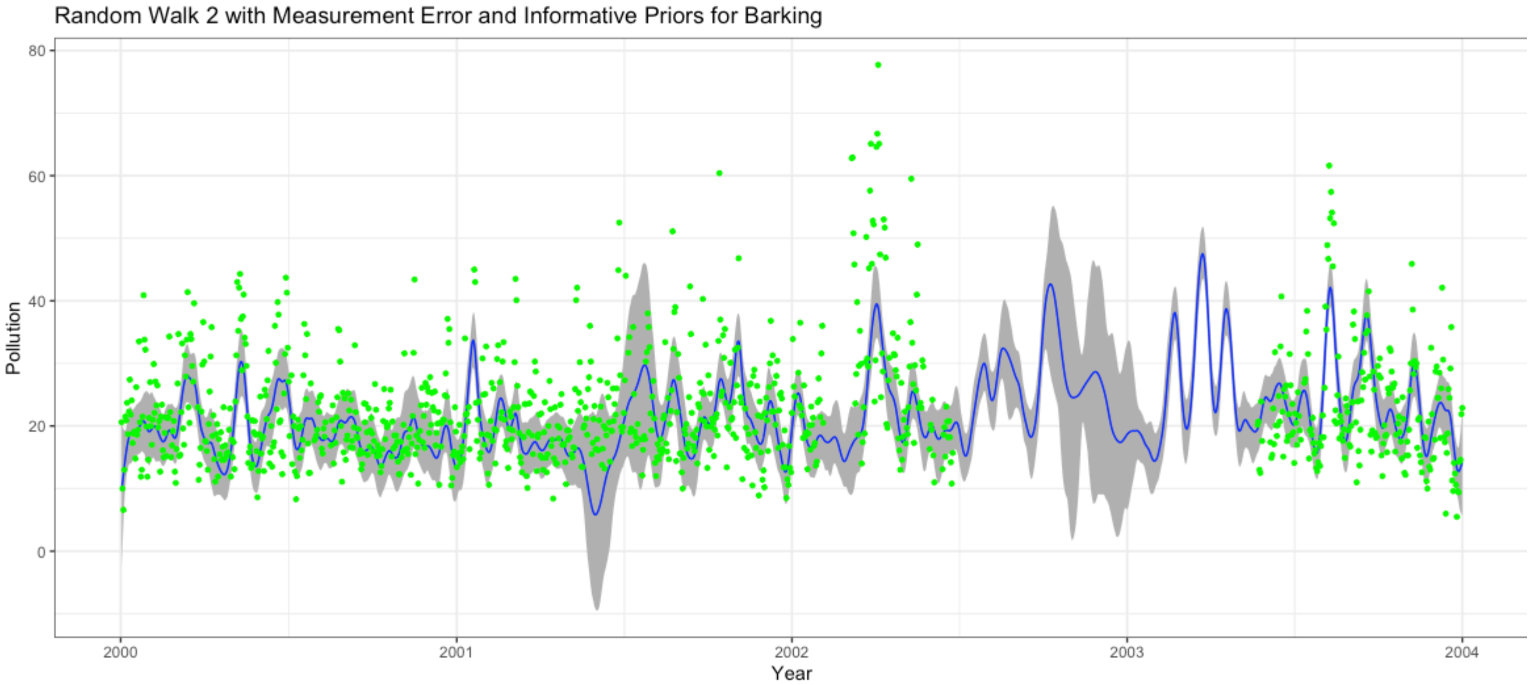


Figure 22



Figures 20 and 21 show the same as Figures 17 and 18, but for RW2. Using informative priors for RW2 results in Figure 22, which also looks fairly similar to the non informative priors RW2 in Figure 16. The periods of missing data differ between the models to a greater extent than they did for RW1, and so the credible intervals are also more different. It could be argued the missing data blue line in Figure 22 is perhaps more expected than that in Figure 16, based on the data points. In terms of convergence, as seen before, the traceplots do not converge, the same as for RW2 non informative priors for Barking and Bloomsbury. And again, this lack of convergence is seen in the gelman diagnostic values and Rhat values in the model summary.

Daanish Ahsan
Advanced Topics in Statistics
Candidate Number: 124070

Code Appendix

```
#Advanced Topics Assignment 2

library(ggplot2)
library(tidyverse)
library(fivethirtyeight)
library(VIM)
library(Amelia)
library(rgdal)
library(coda)
library(rjags)
library(R2jags)
library(coda)
library(dvmisc)
library(kableExtra)

set.seed(1000)

London <- read.csv('London_Pollution.csv')

summary(London)

#Bloomsbury pollution minimum is 5.5, average of 22.09, maximum of 77.7
#427 NA values

#Barking minimum is 3.8, average of 21.53, maximum of 71.8
#232 NA values

Bloomsbury<-  London[,1:2]
summary(Bloomsbury)
y2k <- Bloomsbury[1:366,]
y2001 <- Bloomsbury[367:731,]
y2002 <- Bloomsbury[732:1096,]
y2003 <- Bloomsbury[1097:1461,]
y2004 <- Bloomsbury[1462:1827,]

sum(is.na(y2k$Bloomsbury))
#12 NA values in 2000 in Bloomsbury
sum(is.na(y2001$Bloomsbury))
#12 NA values in 2001
sum(is.na(y2002$Bloomsbury))
#240 NA values in 2002
sum(is.na(y2003$Bloomsbury))
#154 NA values in 2003
sum(is.na(y2004$Bloomsbury))
#9 NA values in 2004

Barking<-  London[,2:3]
summary(Barking)
y2k <- Barking[1:366,]
```

```
y2001 <- Barking[367:731,]
y2002 <- Barking[732:1096,]
y2003 <- Barking[1097:1461,]
y2004 <- Barking[1462:1827,]

sum(is.na(y2k$Barking))
#3 NA values in 2000
sum(is.na(y2001$Barking))
#104 NA values in 2001
sum(is.na(y2002$Barking))
#113 NA values in 2002
sum(is.na(y2003$Barking))
#11 NA values in 2003
sum(is.na(y2004$Barking))
#1 NA value in 2004

#missingness for both increases from 2000 to peak in 2002, falls in 2003
#and then falls dramatically in 2004

London <- London[,2:3]

matrixplot(London)

spineMiss(London)

missmap(London)

#2

ggplot(Bloomsbury, aes(x = c(1:1827), y = Bloomsbury, col='RED')) +
 geom_line(group=1) + theme_bw()+scale_x_continuous(breaks = c(0, 366, 731,
 1096, 1461), labels = c('2000', '2001', '2002', '2003', '2004')) +
  ggtitle("Bloomsbury") +
   annotate('text', x=1096, y=25, label='MISSING DATA', size=7)+
   annotate('text', x=785, y=25, label='MISSING DATA', angle = 90, size=3)+
   labs(y='Pollution', x='Year') +theme(legend.position = 'none')

ggplot() + geom_line(Barking, mapping=aes(x = c(1:1827), y = Barking),
 col='purple') + theme_bw()+scale_x_continuous(breaks = c(0, 366, 731,
 1096, 1461), labels = c('2000', '2001', '2002', '2003', '2004'))+
 ggtitle("Barking") +
   labs(y='Pollution', x='Year') +
   annotate('text', x=550, y=25, label='MISSING DATA', angle=90, size=5)+
   annotate('text', x=1050, y=25, label='MISSING DATA', angle = 90, size=5)+
   theme(legend.position = 'none')

#3

London <- readOGR(dsn = '.', layer = 'London')

plot(London) #map of London
```

```
par(mfrow=c(1,1))
points(548030, 183363, col='red', pch=19, cex=2) #Barking is red
points(530123, 182014, col='blue', pch=19, cex=2) #Bloomsbury is blue
#Bloomsbury has more missing values, but higher average pollution, higher
 minimum pollution,
#and higher maximum pollution compared to Barking.
#This may be bc Bloomsbury is a tourism heavy area in the West End
#so has more air pollution

#4, 5, 6

#Model 1 corrected

BLOOM <- Bloomsbury[1:1461,]
jags.data <- list(Bloomsbury=BLOOM$Bloomsbury,N=1461)
jags.param <- c("Bloomsbury","sigma.w2","tau.w","sigma.v2","tau.v",
                "B.pred", "MSErw1")

#'MSErw1'

# Initial values
miss.v <- is.na(jags.data)
b.init1 <- rep(1,times=1461)
b.init1[miss.v==FALSE] <- NA
b.init1[miss.v==TRUE] <- 22
b.init2 <- rep(1,times=1461)
b.init2[miss.v==FALSE] <- NA
b.init2[miss.v==TRUE] <- 20
B.pred.inits1 = rep(20, 1461)
B.pred.inits2 = rep(20, 1461)
inits1 <- list( "tau.w" = 1, "tau.v" = 1,"Bloomsbury" =
                  b.init1, "B.pred" = B.pred.inits1)
inits2 <- list( "tau.w" = 1, "tau.v" =
                  1,"Bloomsbury"=b.init2 , "B.pred" = B.pred.inits2)
jags.inits <- list(inits1, inits2)

potato <- function() {
  tau.w ~ dgamma(1,0.01)
  sigma.w2 <- 1/tau.w
  tau.v ~ dgamma(1,0.01)
  sigma.v2 <- 1/tau.v


  B.pred[1] ~ dnorm(0, 1.0E-3)
  for(i in 2:N) {
    Bloomsbury[i] ~ dnorm(B.pred[i],tau.v)
    B.pred[i] ~ dnorm(B.pred[i-1], tau.w)
  }
  MSErw1 <- sqrt(sum((B.pred - Bloomsbury)^2)/N)
}
```

```
mod.rw.intercept = jags(jags.data, parameters.to.save = jags.param, inits =
 jags.inits,
                        n.chains = 2, model.file = potato, n.burnin = 5000,
 n.thin = 1,
                        n.iter = 10000, DIC = TRUE)

#MSE

mod.rw.intercept$BUGSoutput$mean$MSErw1

#MSE of 4.25

traceplot(mod.rw.intercept)

#gelman and check convergence, needs to be saved as df, multivariate=False

mod.rw.interceptmc <- (as.mcmc(mod.rw.intercept))

gelmanbloomrw1 <- gelman.diag(mod.rw.interceptmc, multivariate = FALSE)

gelmanbloomrw1 <- as.data.frame(gelmanbloomrw1[["psrf"]])

kable(gelmanbloomrw1, 'html') %>%
  cat(., file = 'GelmanRW1.html')

#no convergence, as not all values are below 1.1

#plot of model

p <- as.data.frame(mod.rw.intercept$BUGSoutput$summary)

p <- p[1:1461,]

lower <- p$`2.5%`

upper <- p$`97.5%`

h <- ggplot(mapping=aes(x=c(1:1461),
 y=mod.rw.intercept$BUGSoutput$mean$B.pred))

h + geom_ribbon(aes(ymin = lower, ymax = upper), fill = "grey70") +
  geom_line(mapping=aes(x=c(1:1461),
 y=mod.rw.intercept$BUGSoutput$mean$B.pred), colour='blue') +
 theme_bw()+scale_x_continuous(breaks = c(0, 366, 731, 1096, 1461), labels
 = c('2000', '2001', '2002', '2003', '2004')) +
  labs(y='Pollution', x='Year') +
  geom_point(mapping = aes(x=c(1:1461),
 y=BLOOM$Bloomsbury),size=1,colour='green')+
  ggtitle('Random Walk 1 with Measurement Error for Bloomsbury')+
  theme(legend.position = 'none')
```

```
p <- as.data.frame(mod.rw.intercept$BUGSoutput$summary)

kable(p, 'html') %>%
  cat(., file = 'SummaryBloomRW1.html')

#plot with first 50 days
p <- as.data.frame(mod.rw.intercept$BUGSoutput$summary)

p <- p[1:50,]

lower <- p$`2.5%`

upper <- p$`97.5%`

j <- as.data.frame(mod.rw.intercept$BUGSoutput$mean$B.pred)
j <- j[1:50,]

h <- ggplot(mapping=aes(x=c(1:50), y=j))

h + geom_ribbon(aes(ymin = lower, ymax = upper), fill = "grey70") +
  geom_line(mapping=aes(x=c(1:50), y=j), colour='blue') +
 theme_bw()+scale_x_continuous(breaks = c(0, 31), labels = c('Jan', 'Feb'))
 +
  labs(y='Pollution', x='Year') +
  geom_point(mapping = aes(x=c(1:50),
 y=Bloomsbury$Bloomsbury[1:50]),size=1,colour='green')+
  ggtitle('Random Walk 1 with Measurement Error over first 50 days for
 Bloomsbury')+
  theme(legend.position = 'none')
#7

#correct model 2

BLOOM <- Bloomsbury[1:1461,]
jags.data <- list(Bloomsbury=BLOOM$Bloomsbury,N=1461)
jags.param <- c("Bloomsbury","sigma.w2","tau.w","sigma.v2","tau.v",
                "B.pred", 'MSErw1')

# Initial values
miss.v <- is.na(jags.data)
b.init1 <- rep(1,times=1461)
b.init1[miss.v==FALSE] <- NA
b.init1[miss.v==TRUE] <- 22
b.init2 <- rep(1,times=1461)
b.init2[miss.v==FALSE] <- NA
b.init2[miss.v==TRUE] <- 20
B.pred.inits1 = rep(20, 1461)
B.pred.inits2 = rep(20, 1461)
inits1 <- list( "tau.w" = 1, "tau.v" = 1,"Bloomsbury" =
                  b.init1, "B.pred" = B.pred.inits1)
inits2 <- list( "tau.w" = 1, "tau.v" =
```

```
                  1,"Bloomsbury"=b.init2 , "B.pred" = B.pred.inits2)
jags.inits <- list(inits1, inits2)

potato3<-function() {

  B.pred[1] ~ dnorm(0, 1.0E-3)
  B.pred[2] ~ dnorm(0, 1.0E-3)
  for (i in 3 : N){
    Bloomsbury[i] ~ dnorm(B.pred[i],tau.v)
    B.pred[i] ~ dnorm(2 * B.pred[i-1] - B.pred[i-2],
                      tau.w)
  }
  MSErw1 <- sqrt(sum((B.pred - Bloomsbury)^2)/N)

  # priors
  tau.w ~ dgamma(1,0.01)
  sigma.w2 <- 1/tau.w
  tau.v ~ dgamma(1,0.01)
  sigma.v2 <- 1/tau.v

}

mod.rw.intercept = jags(jags.data, parameters.to.save = jags.param, inits =
 jags.inits,
                        n.chains = 2, model.file = potato3, n.burnin =
 5000, n.thin = 1,
                        n.iter = 10000, DIC = TRUE)

traceplot(mod.rw.intercept)

mod.rw.intercept$BUGSoutput$mean$MSErw1

#MSE is 7.75.

#gelman and check convergence, needs to be saved as df, multivariate=False

mod.rw.interceptmc <- (as.mcmc(mod.rw.intercept))

gelmanbloomrw2 <- gelman.diag(mod.rw.interceptmc, multivariate = FALSE)

gelmanbloomrw2 <- as.data.frame(gelmanbloomrw2[["psrf"]])

kable(gelmanbloomrw2, 'html') %>%
  cat(., file = 'GelmanRW2.html')

#model does not converge

#plot of model

p <- as.data.frame(mod.rw.intercept$BUGSoutput$summary)
```

```
p <- p[1:1461,]

lower <- p$`2.5%`

upper <- p$`97.5%`

h <- ggplot(mapping=aes(x=c(1:1461),
 y=mod.rw.intercept$BUGSoutput$mean$B.pred))

h + geom_ribbon(aes(ymin = lower, ymax = upper), fill = "grey70") +
  geom_line(mapping=aes(x=c(1:1461),
 y=mod.rw.intercept$BUGSoutput$mean$B.pred), colour='blue') +
 theme_bw()+scale_x_continuous(breaks = c(0, 366, 731, 1096, 1461), labels
 = c('2000', '2001', '2002', '2003', '2004')) +
  labs(y='Pollution', x='Year') +
  geom_point(mapping = aes(x=c(1:1461),
 y=BLOOM$Bloomsbury),size=1,colour='green')+
  ggtitle('Random Walk 2 with Measurement Error for Bloomsbury')+
  theme(legend.position = 'none')

p <- as.data.frame(mod.rw.intercept$BUGSoutput$summary)

kable(p, 'html') %>%
  cat(., file = 'SummaryBloomRW2.html')

#plot with first 50 days
p <- as.data.frame(mod.rw.intercept$BUGSoutput$summary)

p <- p[1:50,]

lower <- p$`2.5%`

upper <- p$`97.5%`

j <- as.data.frame(mod.rw.intercept$BUGSoutput$mean$B.pred)
j <- j[1:50,]

h <- ggplot(mapping=aes(x=c(1:50), y=j))

h + geom_ribbon(aes(ymin = lower, ymax = upper), fill = "grey70") +
  geom_line(mapping=aes(x=c(1:50), y=j), colour='blue') +
 theme_bw()+scale_x_continuous(breaks = c(0, 31), labels = c('Jan', 'Feb'))
 +
  labs(y='Pollution', x='Year') +
  geom_point(mapping = aes(x=c(1:50),
 y=Bloomsbury$Bloomsbury[1:50]),size=1,colour='green')+
  ggtitle('Random Walk 2 with Measurement Error over first 50 days for
 Bloomsbury')+
  theme(legend.position = 'none')

#first 50 days of each model shows that rw 2 has a much larger smoothing
```

```
 effect than rw1

#8, model predictions

#RW 1 prediction, need to double check this

BLOOM <- Bloomsbury[1:1461,]
jags.data <- list(Bloomsbury=c(BLOOM$Bloomsbury, NA, NA, NA, NA, NA, NA,
 NA), N = (1461 + 7))
jags.param <- c("Bloomsbury","sigma.w2","tau.w","sigma.v2","tau.v",
                "B.pred")

# Initial values
miss.v <- is.na(jags.data)
b.init1 <- rep(1,times=1468)
b.init1[miss.v==FALSE] <- NA
b.init1[miss.v==TRUE] <- 22
b.init2 <- rep(1,times=1468)
b.init2[miss.v==FALSE] <- NA
b.init2[miss.v==TRUE] <- 20
B.pred.inits1 = rep(20, 1468)
B.pred.inits2 = rep(20, 1468)
inits1 <- list( "tau.w" = 1, "tau.v" = 1,"Bloomsbury" =
                  b.init1, "B.pred" = B.pred.inits1)
inits2 <- list( "tau.w" = 1, "tau.v" =
                  1,"Bloomsbury"=b.init2 , "B.pred" = B.pred.inits2)
jags.inits <- list(inits1, inits2)

potato <- function() {
  tau.w ~ dgamma(1,0.01)
  sigma.w2 <- 1/tau.w
  tau.v ~ dgamma(1,0.01)
  sigma.v2 <- 1/tau.v


  B.pred[1] ~ dnorm(0, 1.0E-3)
  for(i in 2:N) {
    Bloomsbury[i] ~ dnorm(B.pred[i],tau.v)
    B.pred[i] ~ dnorm(B.pred[i-1], tau.w)
  }
}

jags.param <- c("Bloomsbury","sigma.w2","tau.w","sigma.v2","tau.v",
                "B.pred")
mod_ss_forecast = jags(jags.data, parameters.to.save = jags.param, inits =
 jags.inits,
                       model.file = potato, n.chains = 2, n.burnin = 5000,
 n.thin = 1,
                       n.iter = 10000, DIC = TRUE)

oh <- as.data.frame(mod_ss_forecast$BUGSoutput$mean$B.pred)
```

```
#Predicted against actual

#RW1

h <- ggplot(mapping=aes(x=c(1462:1468),
 y=oh$`mod_ss_forecast$BUGSoutput$mean$B.pred`))

h + geom_line(mapping=aes(x=c(1462:1468),
 y=oh$`mod_ss_forecast$BUGSoutput$mean$B.pred`[1462:1468]), colour='blue')
 + theme_bw()+scale_x_continuous(breaks = c(1462, 1463, 1464, 1465, 1466,
 1467, 1468), labels = c("01/01", "02/01", "03/01", "04/01", "05/01",
 "06/01", "07/01")) +
  labs(y='Pollution', x='First week of 2004') +
  geom_point(mapping = aes(x=c(1462:1468),
 y=Bloomsbury$Bloomsbury[1462:1468]),size=1,colour='green')+
  ggtitle('Random Walk 1 with Predicted Values against actual values for
 First Week of 2004 for Bloomsbury')+
  theme(legend.position = 'none')

 #RW 2 Prediction

BLOOM <- Bloomsbury[1:1461,]
jags.data <- list(Bloomsbury=c(BLOOM$Bloomsbury, NA, NA, NA, NA, NA, NA,
 NA), N = (1461 + 7))
jags.param <- c("Bloomsbury","sigma.w2","tau.w","sigma.v2","tau.v",
                "B.pred")

# Initial values
miss.v <- is.na(jags.data)
b.init1 <- rep(1,times=1468)
b.init1[miss.v==FALSE] <- NA
b.init1[miss.v==TRUE] <- 22
b.init2 <- rep(1,times=1468)
b.init2[miss.v==FALSE] <- NA
b.init2[miss.v==TRUE] <- 20
B.pred.inits1 = rep(20, 1468)
B.pred.inits2 = rep(20, 1468)
inits1 <- list( "tau.w" = 1, "tau.v" = 1,"Bloomsbury" =
                  b.init1, "B.pred" = B.pred.inits1)
inits2 <- list( "tau.w" = 1, "tau.v" =
                  1,"Bloomsbury"=b.init2 , "B.pred" = B.pred.inits2)
jags.inits <- list(inits1, inits2)

potato3<-function() {

  B.pred[1] ~ dnorm(0, 1.0E-3)
  B.pred[2] ~ dnorm(0, 1.0E-3)
  for (i in 3 : N){
    Bloomsbury[i] ~ dnorm(B.pred[i],tau.v)
    B.pred[i] ~ dnorm(2 * B.pred[i-1] - B.pred[i-2],
```

```
                        tau.w)
  }

  # priors
  tau.w ~ dgamma(1,0.01)
  sigma.w2 <- 1/tau.w
  tau.v ~ dgamma(1,0.01)
  sigma.v2 <- 1/tau.v

}


 mod_ss_forecast = jags(jags.data, parameters.to.save = jags.param, inits =
 jags.inits,
                         model.file = potato3, n.chains = 2, n.burnin =
 5000, n.thin = 1,
                         n.iter = 10000, DIC = TRUE)

 oh <- as.data.frame(mod_ss_forecast$BUGSoutput$mean$B.pred)

#Predicted against actual

#RW2

 h <- ggplot(mapping=aes(x=c(1462:1468),
 y=oh$`mod_ss_forecast$BUGSoutput$mean$B.pred`))

 h + geom_line(mapping=aes(x=c(1462:1468),
 y=oh$`mod_ss_forecast$BUGSoutput$mean$B.pred`[1462:1468]), colour='blue')
 + theme_bw()+scale_x_continuous(breaks = c(1462, 1463, 1464, 1465, 1466,
 1467, 1468), labels = c("01/01", "02/01", "03/01", "04/01", "05/01",
 "06/01", "07/01")) +
   labs(y='Pollution', x='First week of 2004') +
   geom_point(mapping = aes(x=c(1462:1468),
 y=Bloomsbury$Bloomsbury[1462:1468]),size=1,colour='green')+
   ggtitle('Random Walk 2 with Predicted Values against actual values for
 First Week of 2004 for Bloomsbury')+
   theme(legend.position = 'none')

#9

#MSE

# Model one has MSE of 4.25
# Model 2 has MSE of 7.75

#Model 1 must be better at forecasting due to lower MSE

#But model 2 seems to have better predictions

#10
```

```
#Repeating RW1 and RW2 but with Barking

 #RW 1

 BARK <- Barking[1:1461,]
 jags.data <- list(Barking=BARK$Barking,N=1461)
 jags.param <- c("Barking","sigma.w2","tau.w","sigma.v2","tau.v",
                 "B.pred")

 miss.v <- is.na(jags.data)
 b.init1 <- rep(1,times=1461)
 b.init1[miss.v==FALSE] <- NA
 b.init1[miss.v==TRUE] <- 22
 b.init2 <- rep(1,times=1461)
 b.init2[miss.v==FALSE] <- NA
 b.init2[miss.v==TRUE] <- 20
 B.pred.inits1 = rep(20, 1461)
 B.pred.inits2 = rep(20, 1461)
 inits1 <- list( "tau.w" = 1, "tau.v" = 1,"Barking" =
                   b.init1, "B.pred" = B.pred.inits1)
 inits2 <- list( "tau.w" = 1, "tau.v" =
                   1,"Barking"=b.init2 , "B.pred" = B.pred.inits2)
 jags.inits <- list(inits1, inits2)

 potato <- function() {
   tau.w ~ dgamma(1,0.01)
   sigma.w2 <- 1/tau.w
   tau.v ~ dgamma(1,0.01)
   sigma.v2 <- 1/tau.v

   B.pred[1] ~ dnorm(0, 1.0E-3)
   for(i in 2:N) {
     Barking[i] ~ dnorm(B.pred[i],tau.v)
     B.pred[i] ~ dnorm(B.pred[i-1], tau.w)
   }
 }

 mod.rw.intercept = jags(jags.data, parameters.to.save = jags.param, inits
= jags.inits,
                        n.chains = 2, model.file = potato, n.burnin =
5000, n.thin = 1,
                        n.iter = 10000, DIC = TRUE)

 traceplot(mod.rw.intercept)

 #gelman and check convergence, needs to be saved as df, multivariate=False

mod.rw.interceptmc <- as.mcmc(mod.rw.intercept)

gelmanbarkrw1 <- gelman.diag(mod.rw.interceptmc, multivariate = FALSE)
```

```
gelmanbarkrw1 <- as.data.frame(gelmanbarkrw1['psrf'])

kable(gelmanbarkrw1, 'html') %>%
  cat(., file = 'GelmanBarkRW1.html')

#model does not converge

 #plot of model

 p <- as.data.frame(mod.rw.intercept$BUGSoutput$summary)

 p <- p[1:1461,]

 lower <- p$`2.5%`

 upper <- p$`97.5%`

 h <- ggplot(mapping=aes(x=c(1:1461),
y=mod.rw.intercept$BUGSoutput$mean$B.pred))

 h + geom_ribbon(aes(ymin = lower, ymax = upper), fill = "grey70") +
   geom_line(mapping=aes(x=c(1:1461),
y=mod.rw.intercept$BUGSoutput$mean$B.pred), colour='blue') +
theme_bw()+scale_x_continuous(breaks = c(0, 366, 731, 1096, 1461), labels
= c('2000', '2001', '2002', '2003', '2004')) +
   labs(y='Pollution', x='Year') +
   geom_point(mapping = aes(x=c(1:1461),
y=BLOOM$Bloomsbury),size=1,colour='green')+
   ggtitle('Random Walk 1 with Measurement Error for Barking')+
   theme(legend.position = 'none')

 p <- as.data.frame(mod.rw.intercept$BUGSoutput$summary)

 kable(p, 'html') %>%
   cat(., file = 'SummaryBarkRW1.html')

 #RW 2

 BARK <- Barking[1:1461,]
 jags.data <- list(Barking=BARK$Barking,N=1461)
 jags.param <- c("Barking","sigma.w2","tau.w","sigma.v2","tau.v",
                 "B.pred")

 miss.v <- is.na(jags.data)
 b.init1 <- rep(1,times=1461)
 b.init1[miss.v==FALSE] <- NA
 b.init1[miss.v==TRUE] <- 22
 b.init2 <- rep(1,times=1461)
 b.init2[miss.v==FALSE] <- NA
 b.init2[miss.v==TRUE] <- 20
```

```
 B.pred.inits1 = rep(20, 1461)
 B.pred.inits2 = rep(20, 1461)
 inits1 <- list( "tau.w" = 1, "tau.v" = 1,"Barking" =
                  b.init1, "B.pred" = B.pred.inits1)
 inits2 <- list( "tau.w" = 1, "tau.v" =
                  1,"Barking"=b.init2 , "B.pred" = B.pred.inits2)
 jags.inits <- list(inits1, inits2)

 potato3<-function() {

   B.pred[1] ~ dnorm(0, 1.0E-3)
   B.pred[2] ~ dnorm(0, 1.0E-3)
   for (i in 3 : N){
     Barking[i] ~ dnorm(B.pred[i],tau.v)
     B.pred[i] ~ dnorm(2 * B.pred[i-1] - B.pred[i-2],
                        tau.w)
   }
   # priors
   tau.w ~ dgamma(1,0.01)
   sigma.w2 <- 1/tau.w
   tau.v ~ dgamma(1,0.01)
   sigma.v2 <- 1/tau.v
 }

 mod.rw.intercept = jags(jags.data, parameters.to.save = jags.param, inits
= jags.inits,
                        n.chains = 2, model.file = potato3, n.burnin =
5000, n.thin = 1,
                        n.iter = 10000, DIC = TRUE)

 traceplot(mod.rw.intercept)

 #gelman and check convergence, needs to be saved as df, multivariate=False

mod.rw.interceptmc <- as.mcmc(mod.rw.intercept)

gelmanbarkrw2 <- gelman.diag(mod.rw.interceptmc, multivariate = FALSE)

gelmanbarkrw2 <- as.data.frame(gelmanbarkrw2['psrf'])

kable(gelmanbarkrw2, 'html') %>%
  cat(., file = 'GelmanBarkRW2.html')

#model also does not converge...

 #plot of model

 p <- as.data.frame(mod.rw.intercept$BUGSoutput$summary)

 p <- p[1:1461,]
```

```
 lower <- p$`2.5%`

 upper <- p$`97.5%`

 h <- ggplot(mapping=aes(x=c(1:1461),
y=mod.rw.intercept$BUGSoutput$mean$B.pred))

 h + geom_ribbon(aes(ymin = lower, ymax = upper), fill = "grey70") +
   geom_line(mapping=aes(x=c(1:1461),
y=mod.rw.intercept$BUGSoutput$mean$B.pred), colour='blue') +
theme_bw()+scale_x_continuous(breaks = c(0, 366, 731, 1096, 1461), labels
= c('2000', '2001', '2002', '2003', '2004')) +
   labs(y='Pollution', x='Year') +
   geom_point(mapping = aes(x=c(1:1461),
y=BLOOM$Bloomsbury),size=1,colour='green')+
   ggtitle('Random Walk 2 with Measurement Error for Barking')+
   theme(legend.position = 'none')

 p <- as.data.frame(mod.rw.intercept$BUGSoutput$summary)

 kable(p, 'html') %>%
   cat(., file = 'SummaryBarkRW2.html')

#11

#Finding distributions that fit these density plots

 #RW1

 plot(density(mod.rw.intercept[["BUGSoutput"]][["sims.list"]][["tau.w"]]),
      main = "Bloomsbury Model RW 1, Posterior of tau.w", lwd = 3, xlim =
c(0, 0.1))

 x <- seq(0.02, 0.10, length = 1000)
 y <- dnorm(x, mean = 0.0370, sd = 0.0053)
 lines(x, y, type = "l", lwd = 2, col = "red")

 plot(density(mod.rw.intercept[["BUGSoutput"]][["sims.list"]][["tau.v"]]),
      main = "Bloomsbury Model RW 1, Posterior of tau.v", lwd = 3, xlim =
c(0, 0.10))

 x <- seq(0.02, 0.10, length = 1000)
 y <- dnorm(x, mean = 0.058, sd = 0.009)
 lines(x, y, type = "l", lwd = 2, col = "red")

 #RW2

 plot(density(mod.rw.intercept[["BUGSoutput"]][["sims.list"]][["tau.w"]]),
      main = "Bloomsbury Model RW 2, Posterior of tau.w", lwd = 3, xlim =
c(0, 0.1))
```

```
 x <- seq(0, 1, length = 1000)
 y <- dnorm(x, mean = 3.5, sd = 1.25)
 lines(x, y, type = "l", lwd = 2, col = "red")

 plot(density(mod.rw.intercept[["BUGSoutput"]][["sims.list"]][["tau.v"]]),
      main = "Bloomsbury Model RW 2, Posterior of tau.v", lwd = 3, xlim =
c(0, 0.10))

 x <- seq(0.01, 0.028, length = 1000)
 y <- dnorm(x, mean = 0.017, sd =0.001)
 lines(x, y, type = "l", lwd = 2, col = "red")

 #using these distributions as priors in new models for Barking

#RW 1

 BARK <- Barking[1:1461,]
 jags.data <- list(Barking=BARK$Barking,N=1461)
 jags.param <- c("Barking","sigma.w2","tau.w","sigma.v2","tau.v",
                 "B.pred")

 miss.v <- is.na(jags.data)
 b.init1 <- rep(1,times=1461)
 b.init1[miss.v==FALSE] <- NA
 b.init1[miss.v==TRUE] <- 22
 b.init2 <- rep(1,times=1461)
 b.init2[miss.v==FALSE] <- NA
 b.init2[miss.v==TRUE] <- 20
 B.pred.inits1 = rep(20, 1461)
 B.pred.inits2 = rep(20, 1461)
 inits1 <- list( "tau.w" = 1, "tau.v" = 1,"Barking" =
                   b.init1, "B.pred" = B.pred.inits1)
 inits2 <- list( "tau.w" = 1, "tau.v" =
                   1,"Barking"=b.init2 , "B.pred" = B.pred.inits2)
 jags.inits <- list(inits1, inits2)

 potato <- function() {
   tau.w ~ dnorm(0.0370, 0.0053)
   sigma.w2 <- 1/tau.w
   tau.v ~ dnorm(0.058, 0.009)
   sigma.v2 <- 1/tau.v

   B.pred[1] ~ dnorm(0, 1.0E-3)
   for(i in 2:N) {
     Barking[i] ~ dnorm(B.pred[i],tau.v)
     B.pred[i] ~ dnorm(B.pred[i-1], tau.w)
   }
 }

 mod.rw.intercept = jags(jags.data, parameters.to.save = jags.param, inits
= jags.inits,
```

```
                              n.chains = 2, model.file = potato, n.burnin =
5000, n.thin = 1,
                              n.iter = 10000, DIC = TRUE)

 p <- as.data.frame(mod.rw.intercept$BUGSoutput$summary)

 p <- p[1:1461,]

 lower <- p$`2.5%`

 upper <- p$`97.5%`

 h <- ggplot(mapping=aes(x=c(1:1461),
y=mod.rw.intercept$BUGSoutput$mean$B.pred))

 h + geom_ribbon(aes(ymin = lower, ymax = upper), fill = "grey70") +
   geom_line(mapping=aes(x=c(1:1461),
y=mod.rw.intercept$BUGSoutput$mean$B.pred), colour='blue') +
theme_bw()+scale_x_continuous(breaks = c(0, 366, 731, 1096, 1461), labels
= c('2000', '2001', '2002', '2003', '2004')) +
   labs(y='Pollution', x='Year') +
   geom_point(mapping = aes(x=c(1:1461),
y=BLOOM$Bloomsbury),size=1,colour='green')+
   ggtitle('Random Walk 1 with Measurement Error and Informative priors for
Barking')+
   theme(legend.position = 'none')

 p <- as.data.frame(mod.rw.intercept$BUGSoutput$summary)

 kable(p, 'html') %>%
   cat(., file = 'infSummaryBarkRW1.html')

 traceplot(mod.rw.intercept)

 #gelman and check convergence, needs to be saved as df, multivariate=False

 mod.rw.interceptmc <- as.mcmc(mod.rw.intercept)

 infgelmanbarkrw1 <- gelman.diag(mod.rw.interceptmc, multivariate = FALSE)

 infgelmanbarkrw1 <- as.data.frame(infgelmanbarkrw1['psrf'])

 kable(infgelmanbarkrw1, 'html') %>%
   cat(., file = 'infGelmanBarkRW1.html')

#RW 2

 BARK <- Barking[1:1461,]
 jags.data <- list(Barking=BARK$Barking,N=1461)
 jags.param <- c("Barking","sigma.w2","tau.w","sigma.v2","tau.v",
                 "B.pred")
```

```
miss.v <- is.na(jags.data)
b.init1 <- rep(1,times=1461)
b.init1[miss.v==FALSE] <- NA
b.init1[miss.v==TRUE] <- 22
b.init2 <- rep(1,times=1461)
b.init2[miss.v==FALSE] <- NA
b.init2[miss.v==TRUE] <- 20
B.pred.inits1 = rep(20, 1461)
B.pred.inits2 = rep(20, 1461)
inits1 <- list( "tau.w" = 1, "tau.v" = 1,"Barking" =
                  b.init1, "B.pred" = B.pred.inits1)
inits2 <- list( "tau.w" = 1, "tau.v" =
                  1,"Barking"=b.init2 , "B.pred" = B.pred.inits2)
jags.inits <- list(inits1, inits2)

potato3<-function() {

  B.pred[1] ~ dnorm(0, 1.0E-3)
  B.pred[2] ~ dnorm(0, 1.0E-3)
  for (i in 3 : N){
    Barking[i] ~ dnorm(B.pred[i],tau.v)
    B.pred[i] ~ dnorm(2 * B.pred[i-1] - B.pred[i-2],
                      tau.w)
  }
  # priors
  tau.w ~ dnorm(3.5,1.25)
  sigma.w2 <- 1/tau.w
  tau.v ~ dnorm(0.017,0.001)
  sigma.v2 <- 1/tau.v
}

mod.rw.intercept = jags(jags.data, parameters.to.save = jags.param, inits
= jags.inits,
                        n.chains = 2, model.file = potato3, n.burnin =
5000, n.thin = 1,
                        n.iter = 10000, DIC = TRUE)

p <- as.data.frame(mod.rw.intercept$BUGSoutput$summary)

p <- p[1:1461,]

lower <- p$`2.5%`

upper <- p$`97.5%`

h <- ggplot(mapping=aes(x=c(1:1461),
y=mod.rw.intercept$BUGSoutput$mean$B.pred))

h + geom_ribbon(aes(ymin = lower, ymax = upper), fill = "grey70") +
  geom_line(mapping=aes(x=c(1:1461),
```

```
y=mod.rw.intercept$BUGSoutput$mean$B.pred), colour='blue') +
theme_bw()+scale_x_continuous(breaks = c(0, 366, 731, 1096, 1461), labels
= c('2000', '2001', '2002', '2003', '2004')) +
  labs(y='Pollution', x='Year') +
  geom_point(mapping = aes(x=c(1:1461),
y=BLOOM$Bloomsbury),size=1,colour='green')+
  ggtitle('Random Walk 2 with Measurement Error and Informative Priors for
Barking')+
  theme(legend.position = 'none')

p <- as.data.frame(mod.rw.intercept$BUGSoutput$summary)

kable(p, 'html') %>%
  cat(., file = 'infSummaryBarkRW2.html')

traceplot(mod.rw.intercept)

#gelman and check convergence, needs to be saved as df, multivariate=False

mod.rw.interceptmc <- as.mcmc(mod.rw.intercept)

infgelmanbarkrw2 <- gelman.diag(mod.rw.interceptmc, multivariate = FALSE)

infgelmanbarkrw2 <- as.data.frame(infgelmanbarkrw2['psrf'])

kable(infgelmanbarkrw2, 'html') %>%
  cat(., file = 'infgelmanBarkRW2.html')
```