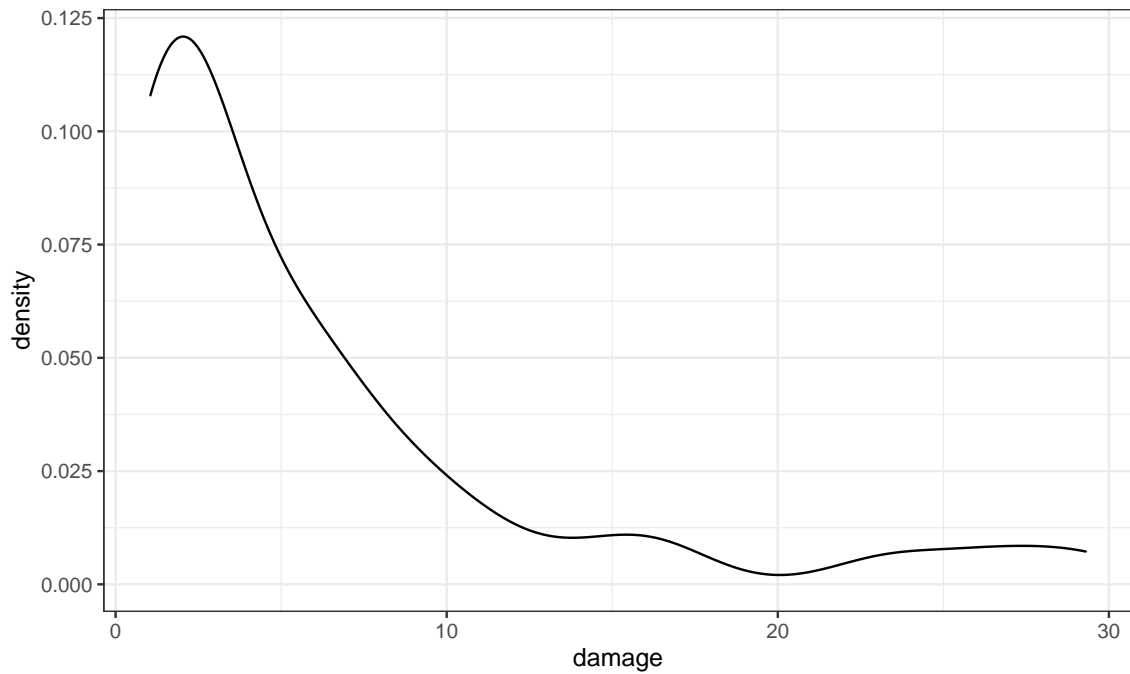


Topics in Advanced Statistical Computing

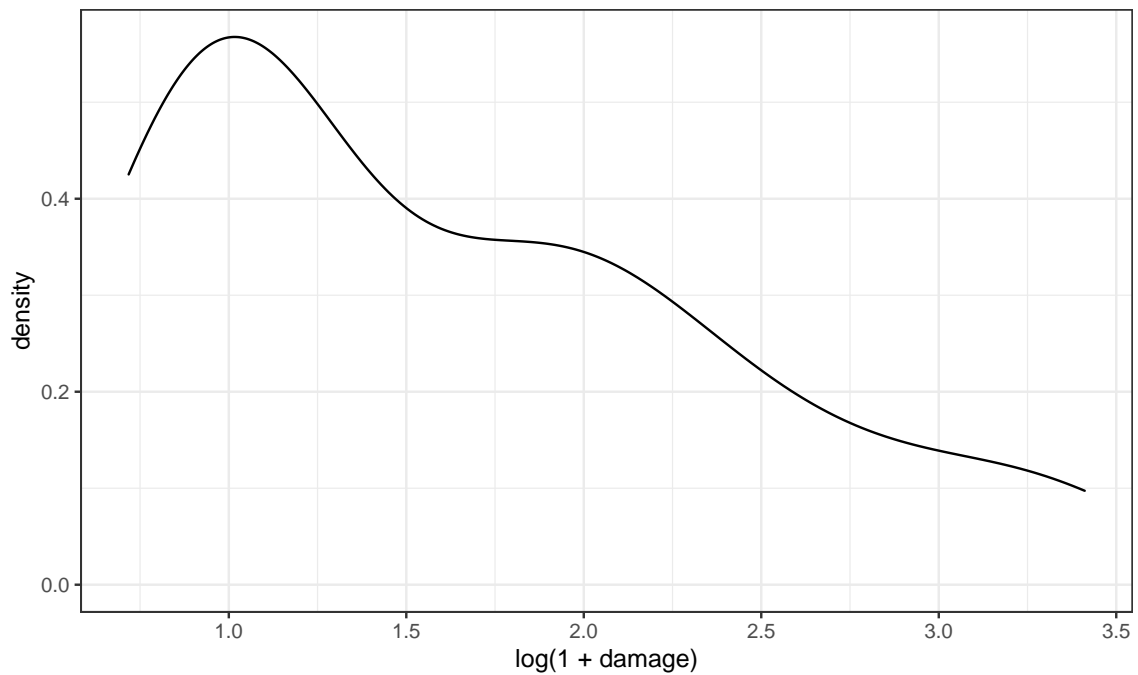
Project 3; Problem 6

Jimmy Kelliher (UNI: jmk2303)

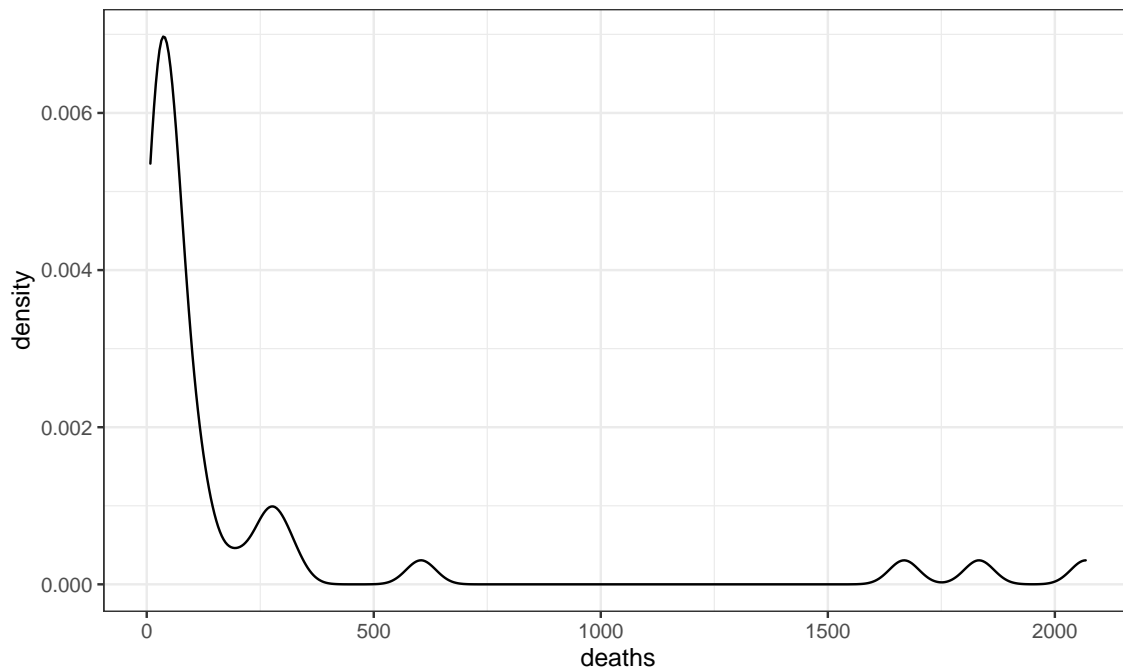
```
hurr <-  
  read_csv("hurricaneoutcome2.csv", show_col_types = FALSE) %>%  
  janitor::clean_names() %>%  
  mutate(damage = as.numeric(str_replace(damage, "\\$", "")))  
  
hurr  
  
## # A tibble: 43 x 14  
##   hurrican_id season damage deaths month nature maxspeed meanspeed maxpressure  
##   <chr>         <dbl> <dbl> <dbl> <chr> <chr>      <dbl>      <dbl>      <dbl>  
## 1 betsy.1965    1965  1.42   81 Sept~ TS        135      75.6      1010  
## 2 camille.1969  1969  1.42  256 Augu~ TS        150      74.0       991  
## 3 agnes.1972    1972  2.1   124 June  TS         75      44.7      1001  
## 4 fifi.1974     1974  1.8  8000 Sept~ TS         95      52.7      1005  
## 5 david.1979    1979  1.54  2068 Sept~ TS        150      82.8      1008  
## 6 frederic.19~  1979  2.3    12 Sept~ TS        115      51.9      1009  
## 7 allen.1980    1980  1.24  269 Augu~ NR        165      97.1      1010  
## 8 alicia.1983   1983  2      21 Augu~ TS        100      49.2      1011  
## 9 elena.1985    1985  1.25    9 Sept~ TS        110      65.4      1012  
## 10 juan.1985    1985  1.5    74 Octo~ TS         75      53.3      1005  
## # ... with 33 more rows, and 5 more variables: meanpressure <dbl>, hours <dbl>,  
## #   total_pop <dbl>, percent_poor <dbl>, percent_usa <dbl>  
  
hurr %>%  
  filter(damage <= quantile(damage, 0.95)) %>%  
  ggplot(aes(x = damage)) +  
  geom_density()
```



```
hurr %>%
  filter(damage <= quantile(damage, 0.95)) %>%
  ggplot(aes(x = log(1 + damage))) +
  geom_density()
```



```
hurr %>%
  filter(deaths <= quantile(deaths, 0.95)) %>%
  ggplot(aes(x = deaths)) +
  geom_density()
```



```
hurr %>% mutate(ratio = deaths / total_pop) %>% arrange(-ratio)
```

```
## # A tibble: 43 x 15
##   hurrican_id season damage deaths month nature maxspeed meanspeed maxpressure
##   <chr>         <dbl> <dbl> <dbl> <chr> <chr>      <dbl>      <dbl>      <dbl>
## 1 mitch.1998    1998   6.2  19325 Octo~ TS        155       63.1      1005
## 2 fifi.1974     1974   1.8   8000 Sept~ TS         95       52.7      1005
## 3 stan.2005     2005   3.96  1668 Octo~ TS         70       39.7      1007
## 4 jeanne.2004   2004   7.66  3035 Sept~ TS        105       59.4      1010
## 5 sandy.2012    2012  71.4   286 Octo~ DS        100       61.6      1006
## 6 david.1979    1979   1.54  2068 Sept~ TS        150       82.8      1008
## 7 alex.2010     2010   1.89   52 June DS         95       51.7      1007
## 8 katrina.2005  2005  125   1833 Augu~ TS        150       74.1      1008
## 9 georges.1998  1998   9.72   604 Sept~ TS        135       75.4      1009
## 10 luis.1995     1995   2.5    19 Sept~ TS        120       86.5      1010
## # ... with 33 more rows, and 6 more variables: meanpressure <dbl>, hours <dbl>,
## #   total_pop <dbl>, percent_poor <dbl>, percent_usa <dbl>, ratio <dbl>
```

```
cutoff <- 1000
```

```
fit_damage <- glm(
  deaths ~ .
  , data = hurr %>% select(-c(hurrican_id, damage)) %>% filter(deaths < cutoff)
  , family = "poisson"
)
```

```
summary(fit_damage)
```

```
##
```

```
## Call:
```

```
## glm(formula = deaths ~ ., family = "poisson", data = hurr %>%
```

```
##   select(-c(hurrican_id, damage)) %>% filter(deaths < cutoff))
```

```
##
```

```
## Deviance Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -11.197  -6.436  -1.631   3.136  29.892
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.545e+02  1.486e+01  10.401 < 2e-16 ***
## season       -1.477e-02  3.621e-03  -4.079 4.52e-05 ***
## monthJuly    -1.472e+00  4.931e-01  -2.986 0.002827 **
## monthJune     3.926e-01  1.062e-01   3.697 0.000218 ***
## monthNovember -3.052e-01  1.648e-01  -1.852 0.063994 .
## monthOctober   8.184e-02  7.900e-02   1.036 0.300230
## monthSeptember 1.905e-01  7.460e-02   2.554 0.010653 *
## natureNR       1.960e-01  1.836e-01   1.067 0.285780
## natureTS      -7.907e-01  8.771e-02  -9.015 < 2e-16 ***
## maxspeed      8.813e-03  1.650e-03   5.341 9.25e-08 ***
## meanspeed     5.057e-03  2.577e-03   1.962 0.049712 *
## maxpressure   -1.225e-01  9.725e-03 -12.596 < 2e-16 ***
## meanpressure   1.051e-03  3.034e-04   3.463 0.000534 ***
## hours         3.753e-03  2.925e-04  12.831 < 2e-16 ***
## total_pop    -1.361e-08  4.491e-08  -0.303 0.761864
## percent_poor   2.622e-02  5.277e-03   4.969 6.72e-07 ***
## percent_usa   -3.731e-03  5.657e-04  -6.595 4.25e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 3840.1  on 36  degrees of freedom
## Residual deviance: 2345.9  on 20  degrees of freedom
## AIC: 2593.7
##
## Number of Fisher Scoring iterations: 6

hurr %>%
  filter(deaths < cutoff) %>%
  mutate(predict = exp(predict(fit_damage))) %>%
  ggplot(aes(x = deaths, y = predict)) +
  geom_point()
```

