

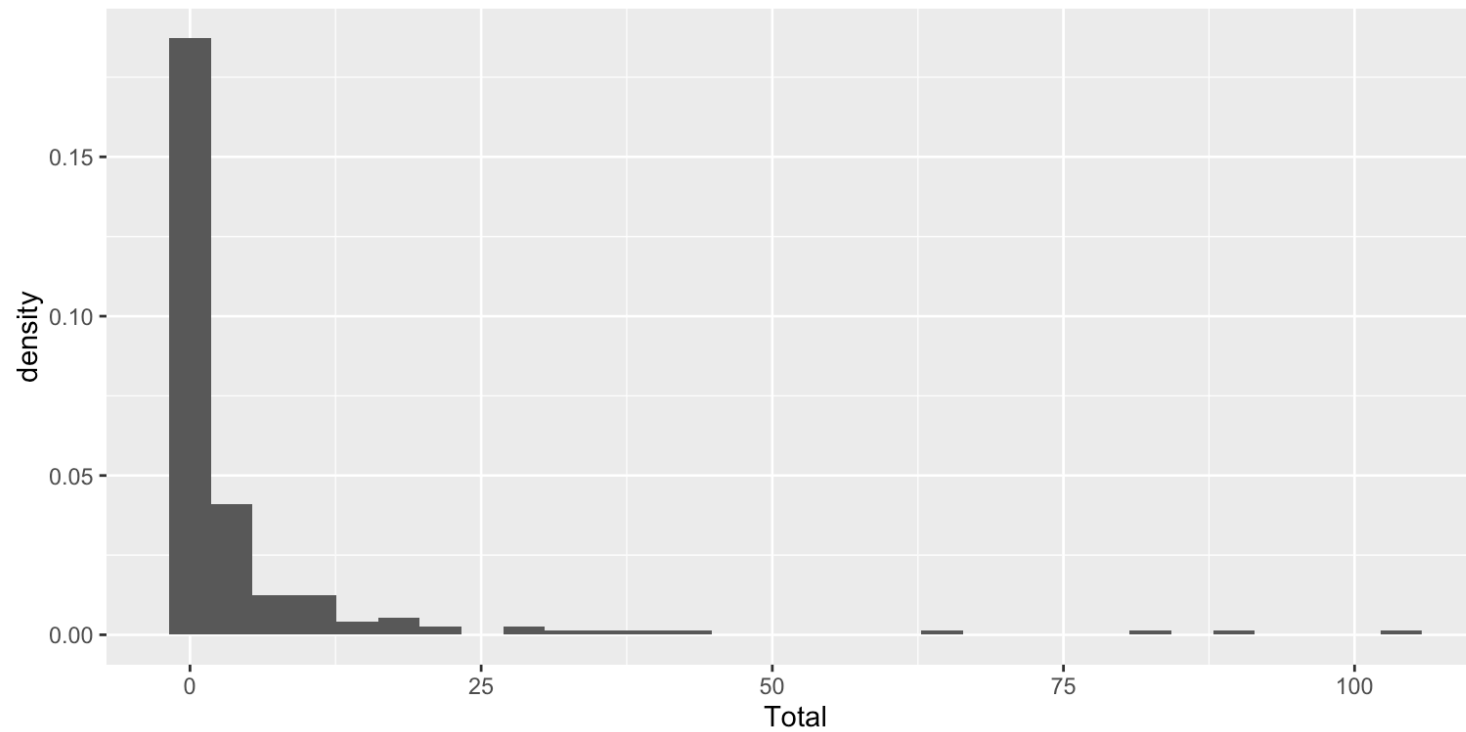
Statistical Methods for Insurance: Fitting Models

Di Cook & Souhaib Ben Taieb, Econometrics and Business Statistics, Monash University
W4.C1

Overview of this class

- Quiz 2
- Fit distribution for olympic medals

Olympic medals, 2012 London



Data

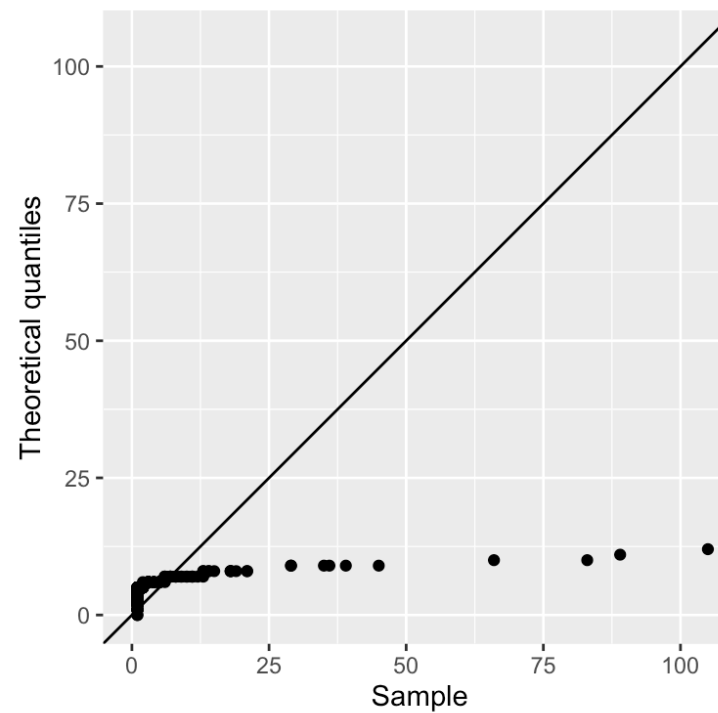
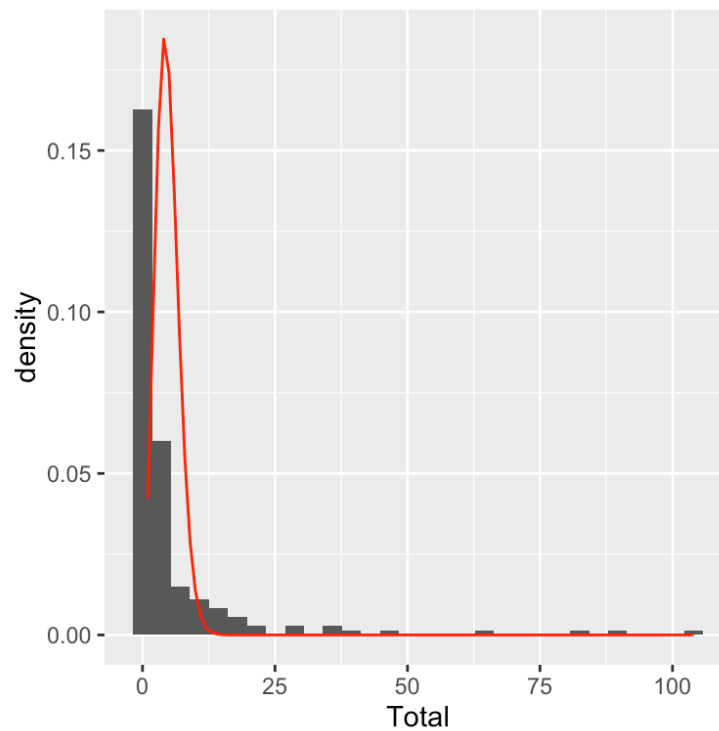
- Extracted from <https://www.olympic.org/london-2012>
- Now it is easier to pull data from [wikipedia](#)
- 204 countries participated, only countries that scored a medal (85) are listed in the medal table

Medal tally

- Examine the distribution of medal counts
- Need to add 119 zeros, to account for participating countries that did not get a medal
- Distribution is right-skewed, heavily, and unimodal
- Use maximum likelihood to estimate parameters for plausible distributions

Fit distribution using Poisson

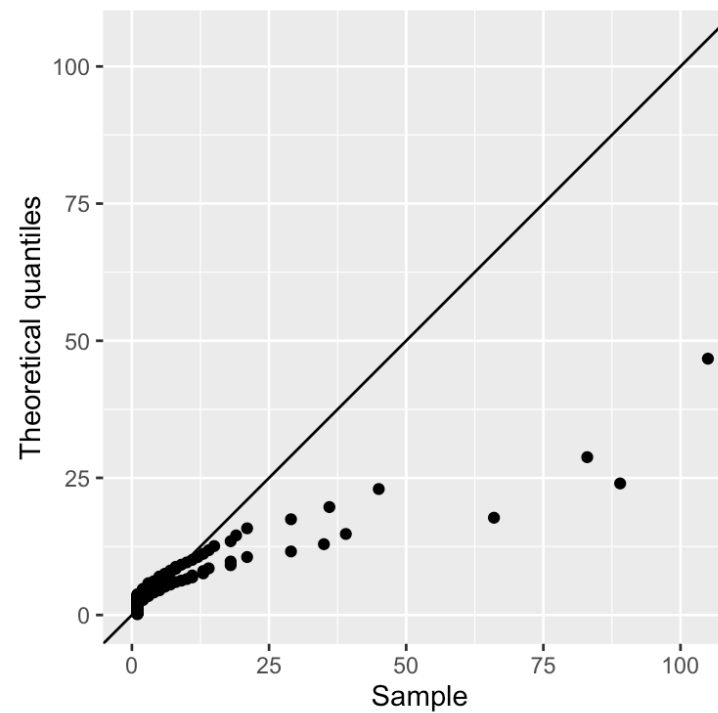
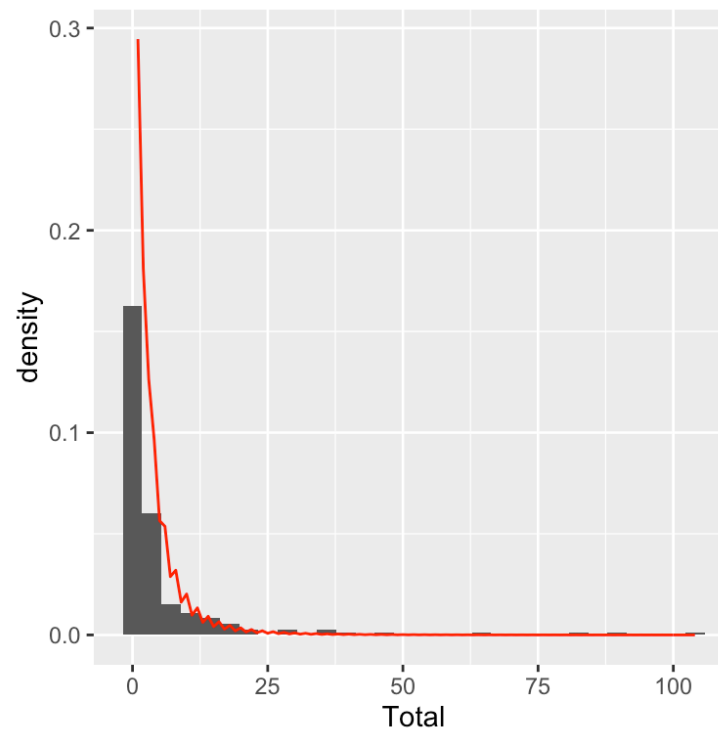
```
#> lambda  
#> 4.72  
#> (0.15)
```



6/20

Try lognormal

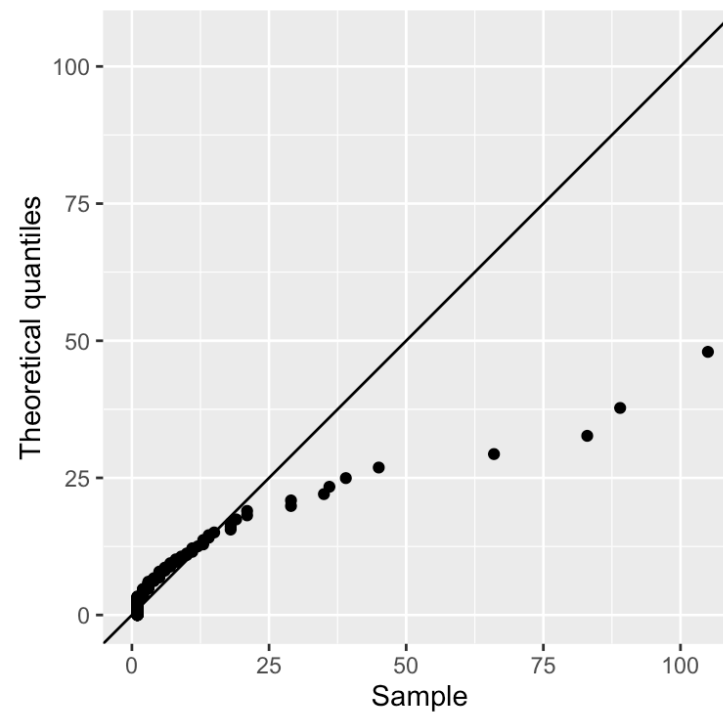
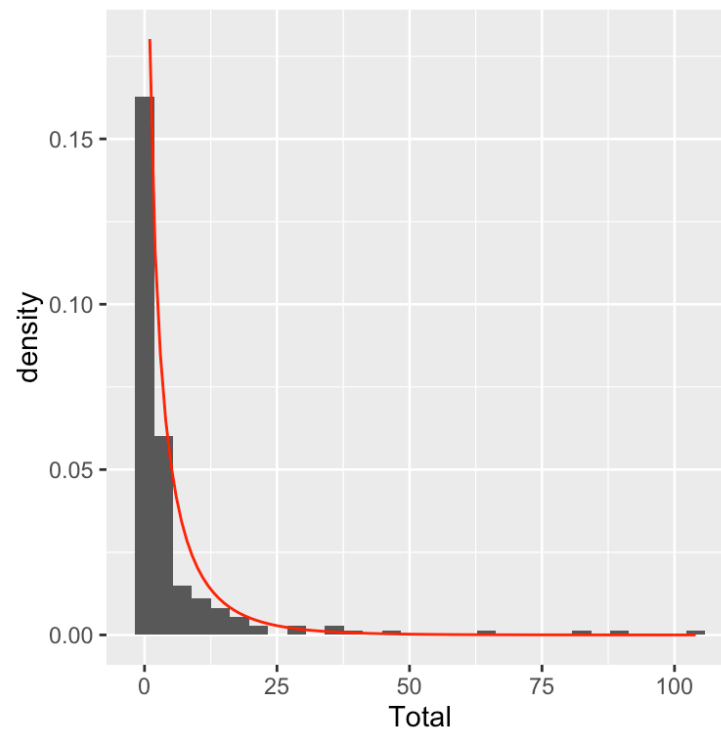
```
#> meanlog sdlog  
#> 0.779 1.137  
#> (0.080) (0.056)
```



7/20

Try weibull

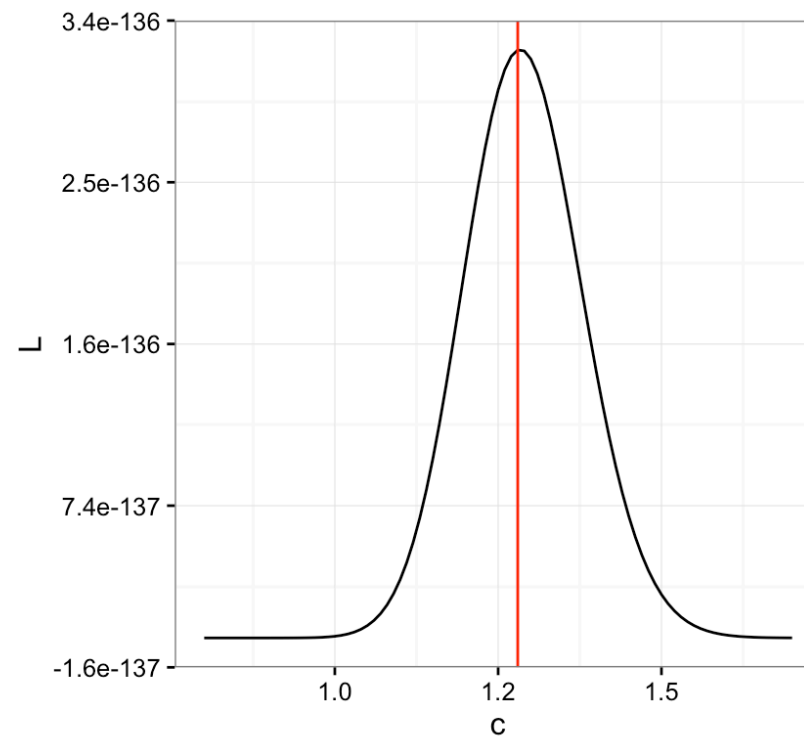
```
#>      shape      scale  
#> 0.707      4.106  
#> (0.033) (0.434)
```



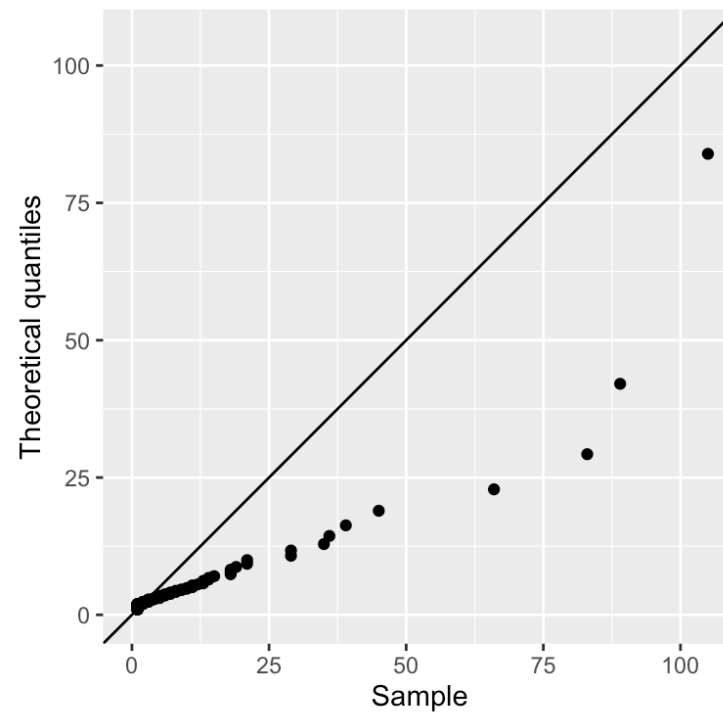
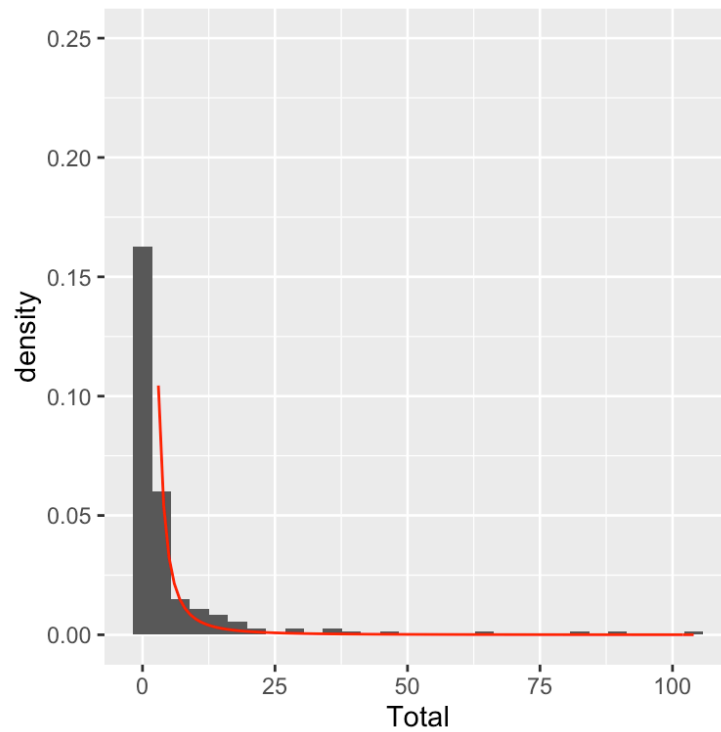
8/20

Try pareto

```
#>      c  
#>  1.28  
#> (0.09)
```

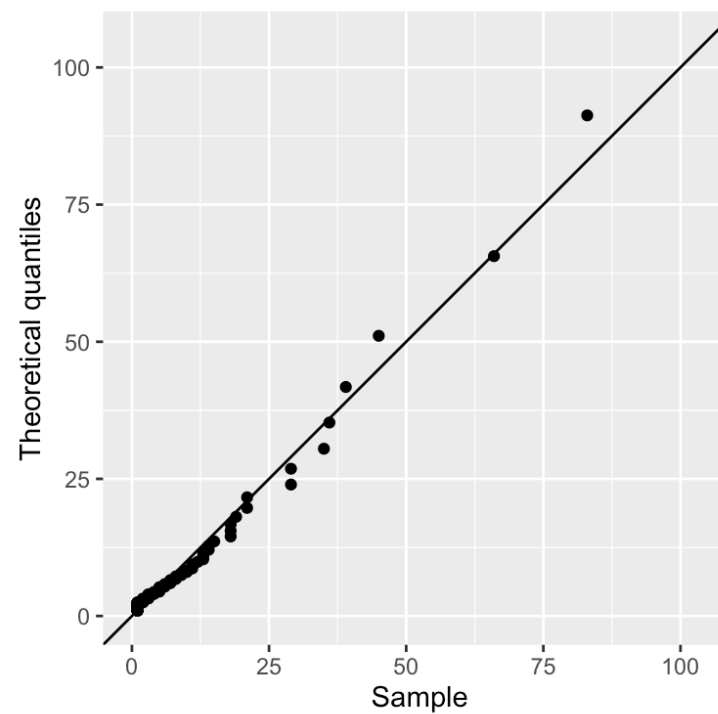
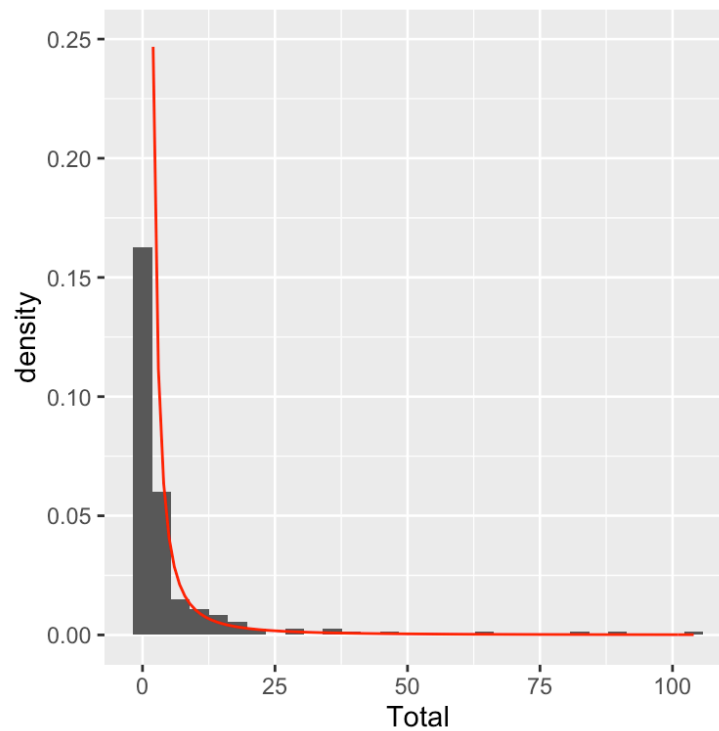


Optimization actually fails



Manually

Actually using $c = 0.96$.



Predict largest medal count

Using this model, what is the probability of observing a tally of more than 50 medals for a country? $P(X > 50)$

```
ppareto <- function(q, c) {  
  if (c<=0) stop("c must be positive > 0")  
  ifelse(q<1, 0, 1-1/q^c)  
}  
1-ppareto(50, 0.96)  
#> [1] 0.023
```

How many would we expect?

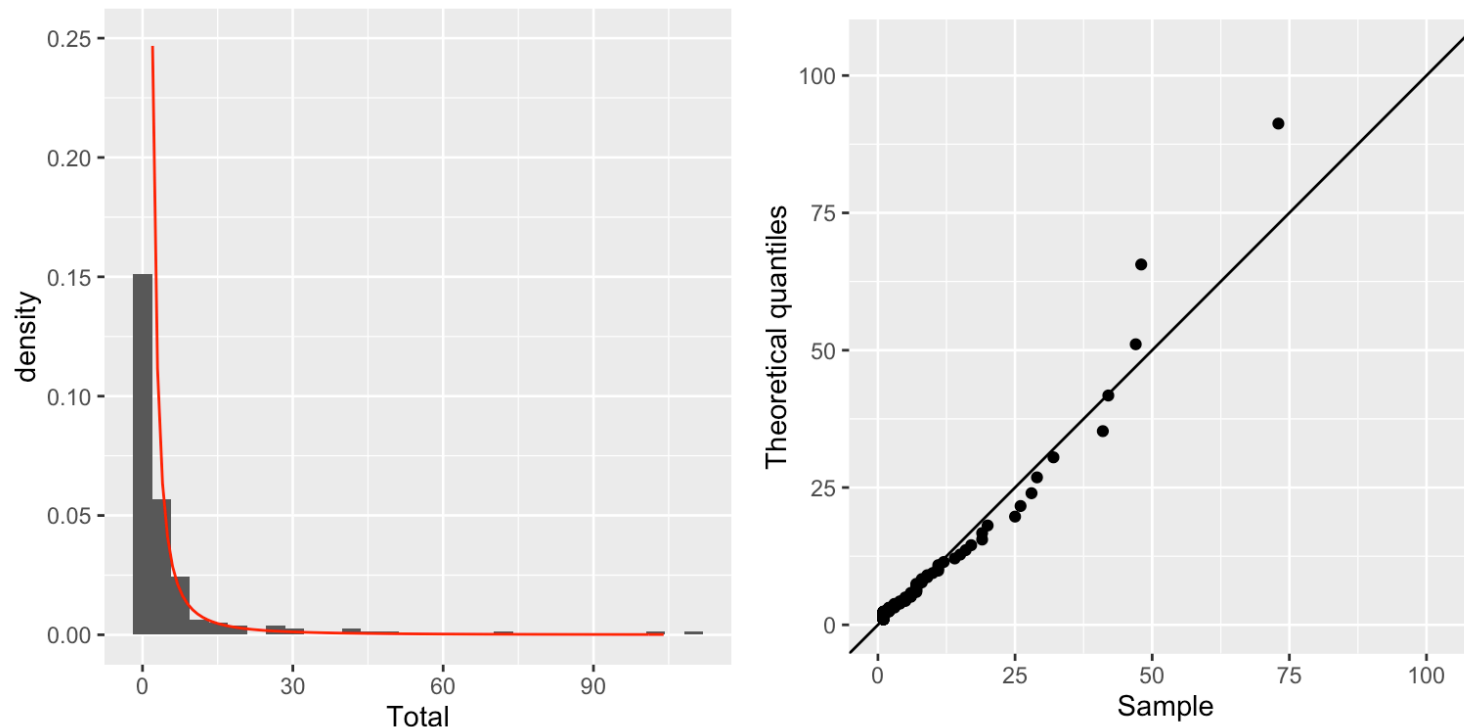
If there are 204 countries, how many of them would we expect to earn more than 50 medals, assuming the *Pareto*(0.96) model?

```
204*(1-ppareto(50, 0.96))  
#> [1] 4.8
```

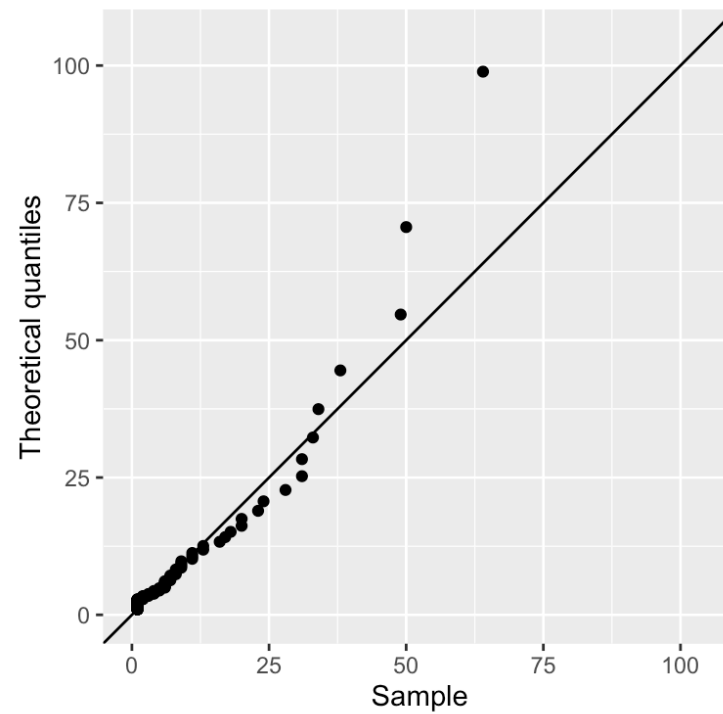
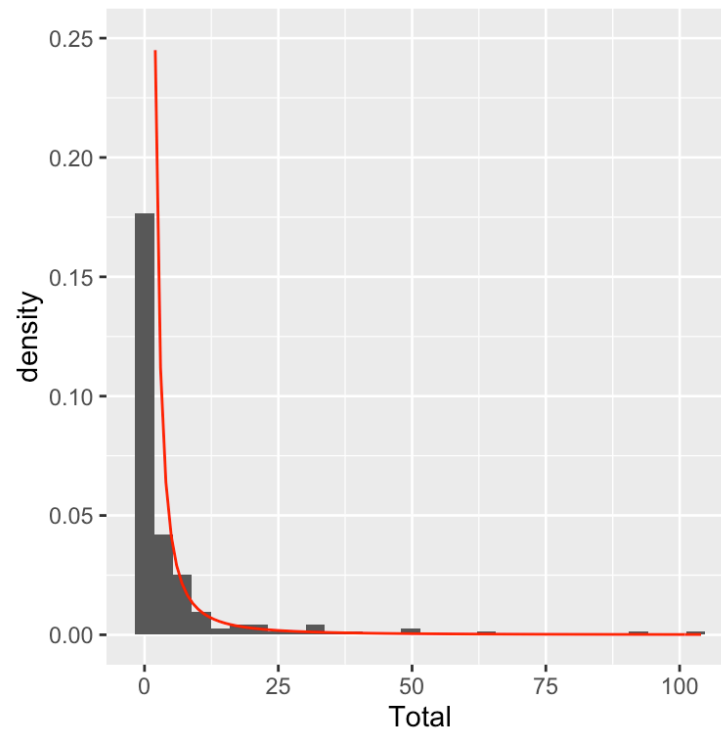
and how does this compare to the observed number?

```
library(dplyr)  
df %>% filter(Total>50)  
#>   Total  
#> 1    65  
#> 2    82  
#> 3    88  
#> 4   104
```

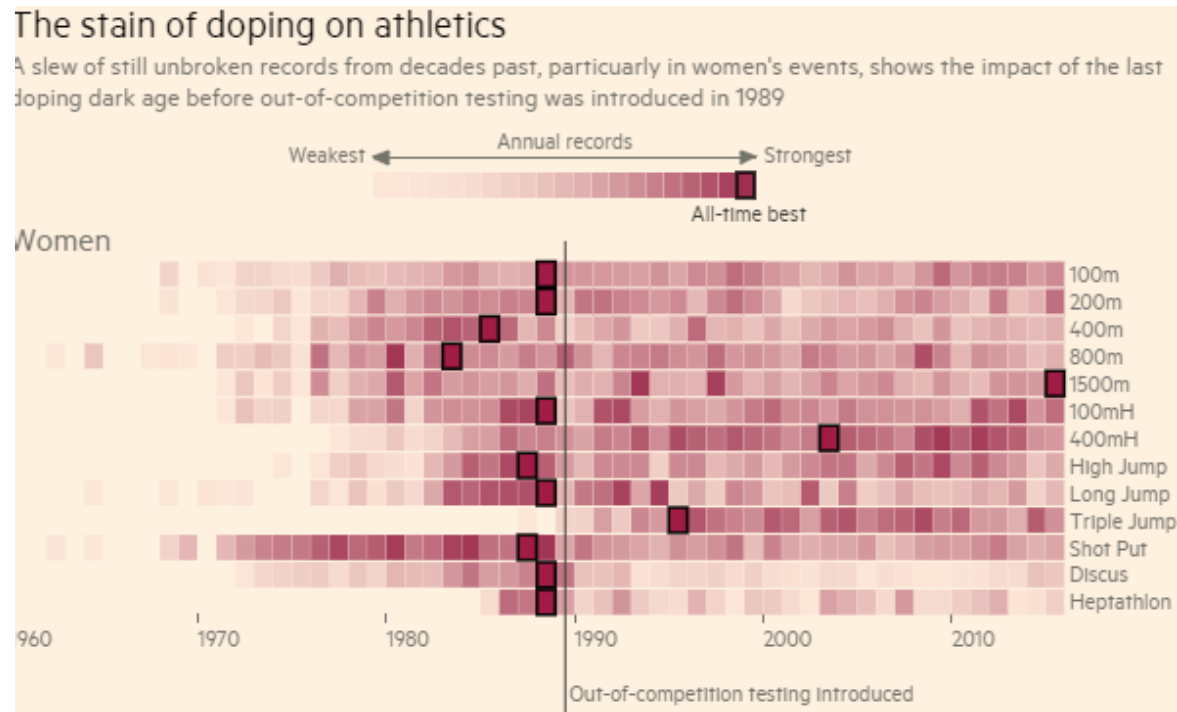
How well does this fit 2008 medal tally?



And 2004?



Doping in sports - finding anomalies



Source: FT research, image extracted from
<http://blogs.ft.com/ftdata/2015/11/16/doping-in-athletics/>

YOUR TURN: How could we improve the model?

- What dependencies are there in the medal tallies?
- What varies among Olympic years?
- What factors might affect the medal counts?

Resources

- [2012 Medal tally](#)
- [2008 Medal tally](#)
- [2004 Medal tally](#)
- <http://blogs.ft.com/ftdata/2015/11/16/doping-in-athletics/>

Share and share alike

This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 United States License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/3.0/us/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.