

<http://courses.had.co.nz>

# Case study

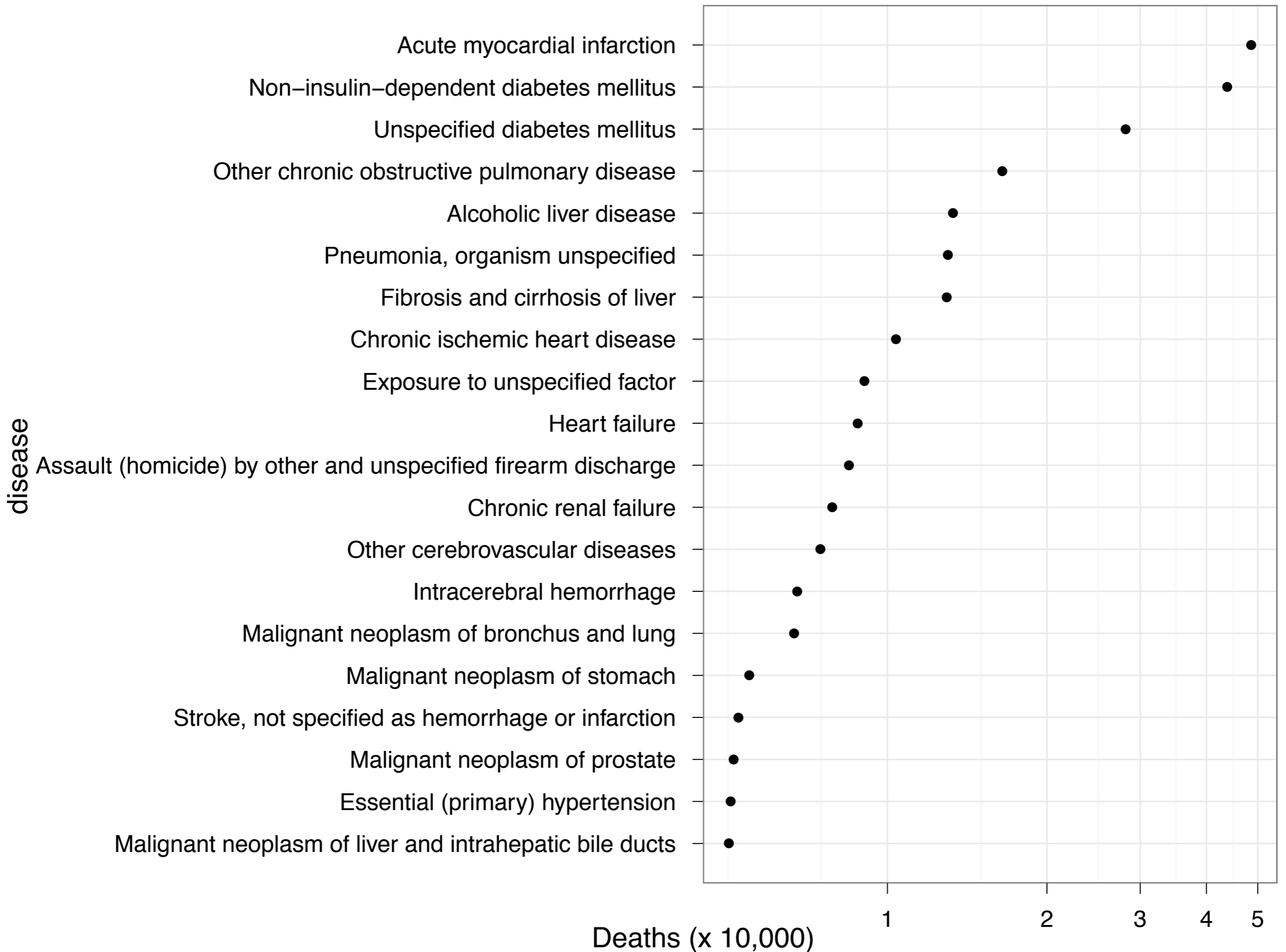
**Hadley Wickham**

Assistant Professor / Dobelman Family Junior Chair  
Department of Statistics / Rice University

**July 2012**



Sunday, July 15, 12





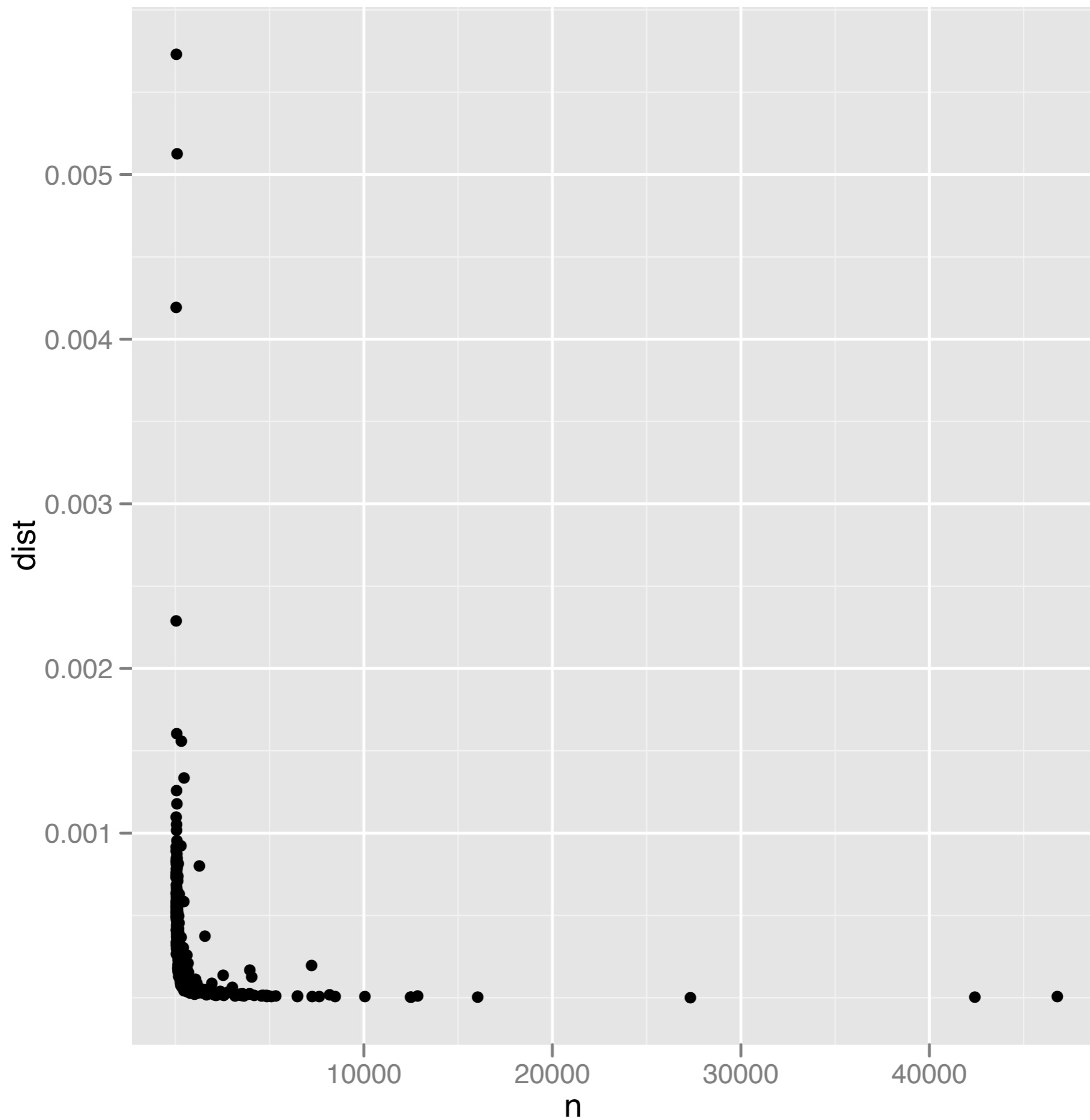
```
hod2 <- count(deaths, c("cod", "hod"))
hod2 <- subset(hod2, !is.na(hod))
hod2 <- join(hod2, codes)
hod2 <- ddply(hod2, "cod", transform,
  prop = freq / sum(freq))

# Compare to overall abundance
overall <- ddply(hod2, "hod", summarise,
  freq_all = sum(freq))
overall <- mutate(overall,
  prop_all = freq_all / sum(freq_all))

hod2 <- join(overall, hod2, by = "hod")
```

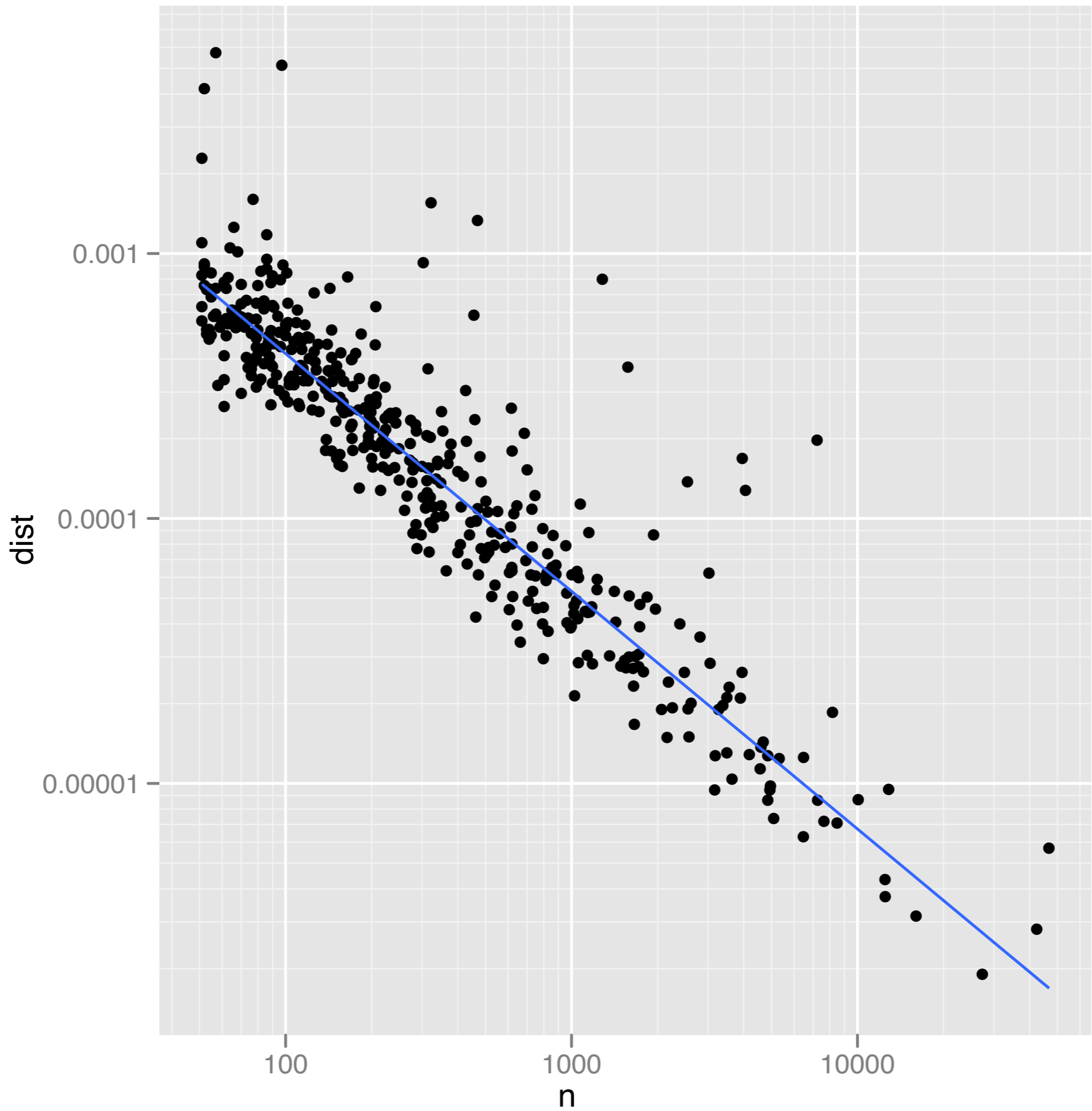
	cod	hod	disease	freq	prop	freq_all	prop_all
1	A01	1	Typhoid and paratyphoid\nfevers	3	0.0577	20430	0.0398
2	A01	2	Typhoid and paratyphoid\nfevers	1	0.0192	18962	0.0369
3	A01	3	Typhoid and paratyphoid\nfevers	4	0.0769	19729	0.0384
4	A01	5	Typhoid and paratyphoid\nfevers	5	0.0962	22126	0.0431
5	A01	6	Typhoid and paratyphoid\nfevers	1	0.0192	23787	0.0463
6	A01	8	Typhoid and paratyphoid\nfevers	1	0.0192	21915	0.0427
7	A01	10	Typhoid and paratyphoid\nfevers	2	0.0385	24321	0.0474
8	A01	11	Typhoid and paratyphoid\nfevers	2	0.0385	23843	0.0465
9	A01	12	Typhoid and paratyphoid\nfevers	1	0.0192	23392	0.0456
10	A01	13	Typhoid and paratyphoid\nfevers	6	0.1154	23284	0.0454
11	A01	14	Typhoid and paratyphoid\nfevers	4	0.0769	23053	0.0449
12	A01	15	Typhoid and paratyphoid\nfevers	5	0.0962	23278	0.0454
13	A01	17	Typhoid and paratyphoid\nfevers	3	0.0577	23625	0.0460
14	A01	18	Typhoid and paratyphoid\nfevers	2	0.0385	24380	0.0475
15	A01	19	Typhoid and paratyphoid\nfevers	3	0.0577	22919	0.0447
16	A01	20	Typhoid and paratyphoid\nfevers	3	0.0577	22926	0.0447
17	A01	21	Typhoid and paratyphoid\nfevers	2	0.0385	20995	0.0409
18	A01	22	Typhoid and paratyphoid\nfevers	3	0.0577	20510	0.0400
19	A01	23	Typhoid and paratyphoid\nfevers	1	0.0192	21446	0.0418

```
devi <- ddply(hod2, "cod", summarise, n = sum(freq),  
  dist = mean((prop - prop_all)^2))  
devi <- subset(devi, n > 50)  
  
qplot(n, dist, data = devi)
```

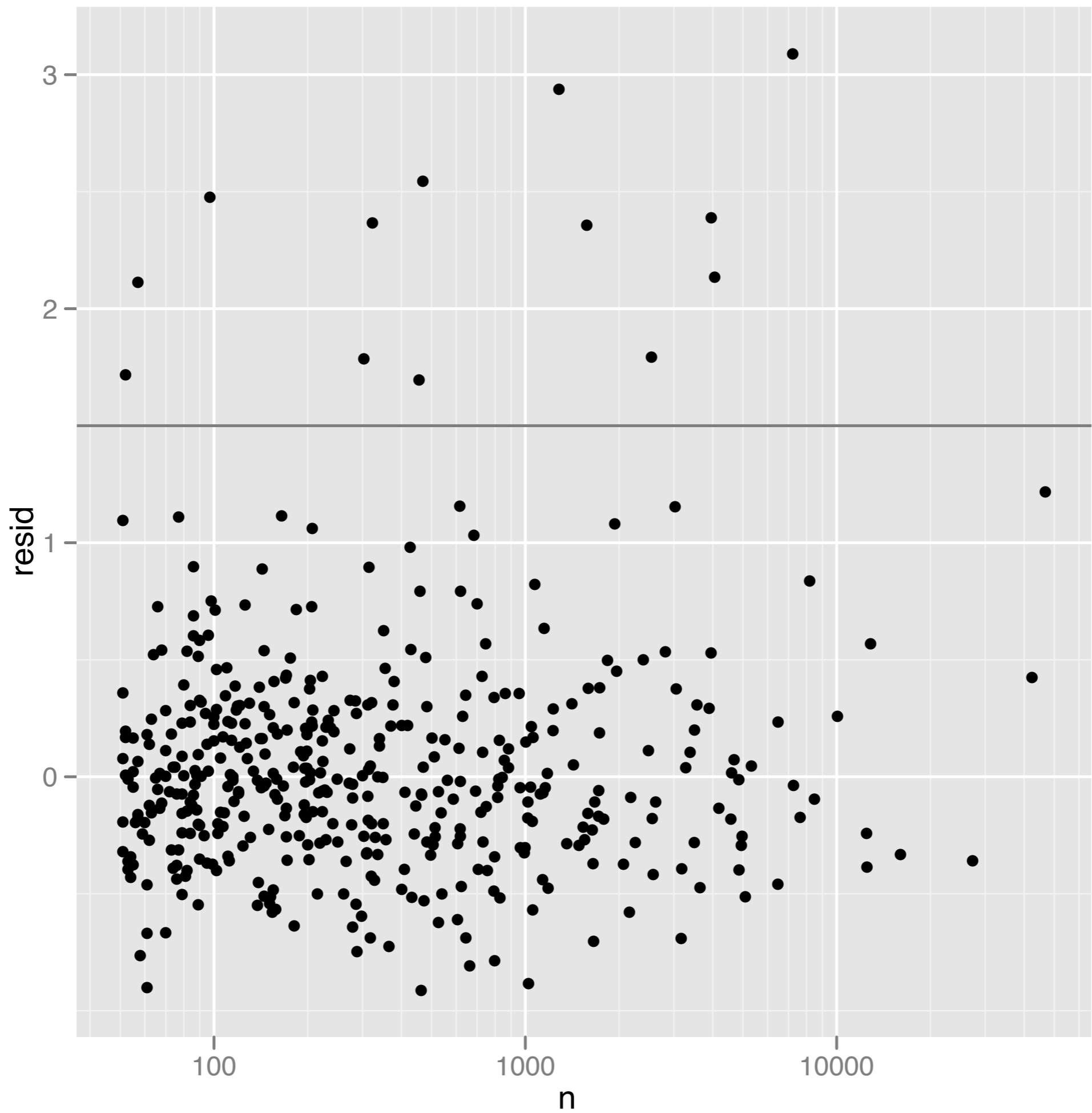


```
qplot(n, dist, data = devi) +  
  geom_smooth(method = "rlm", se = F) +  
  xlog10 +  
  ylog10  
  
xlog10 <- scale_x_log10(  
  breaks = c(100, 1000, 10000),  
  labels = c(100, 1000, 10000),  
  minor_breaks = outer(1:9, 10^(1:5), "*"))  
ylog10 <- scale_y_log10(  
  breaks = 10 ^ -c(3, 4, 5),  
  labels = c("0.001", "0.0001", "0.00001"),  
  minor_breaks = outer(1:9, 10^-(3:6), "*"))
```





```
devi$resid <- resid(rlm(log(dist) ~ log(n),  
  data = devi))  
  
ggplot(devi, aes(n, resid)) +  
  geom_hline(yintercept = 1.5, colour = "grey50") +  
  geom_point() +  
  xlog10
```

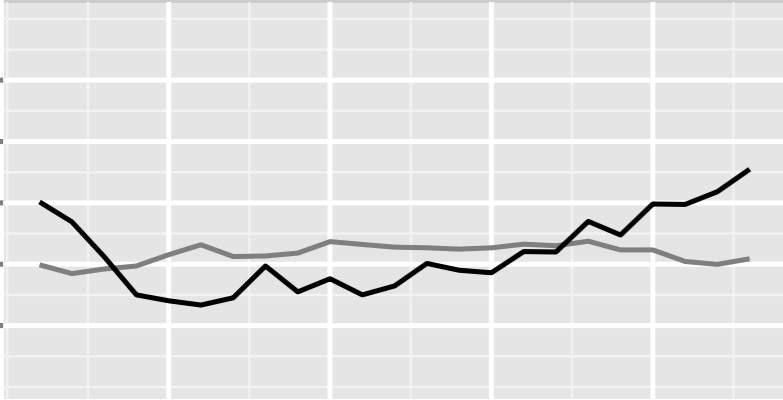


```
unusual <- subset(devi, resid > 1.5)
hod_unusual_big <- match_df(hod2, subset(unusual, n > 350))
hod_unusual_sml <- match_df(hod2, subset(unusual, n <= 350))

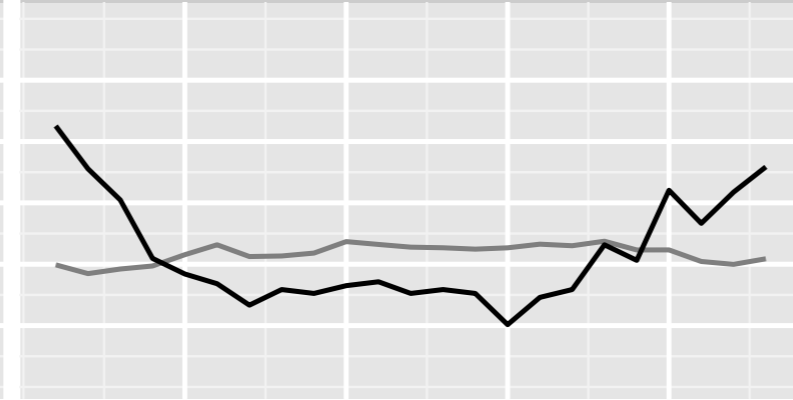
# Visualise unusual causes of death
ggplot(hod_unusual_big, aes(hod, prop)) +
  geom_line(aes(y = prop_all), data = overall, colour = "grey50") +
  geom_line() +
  facet_wrap(~ disease, ncol = 3)
```

Assault (homicide) by other and unspecified firearm discharge

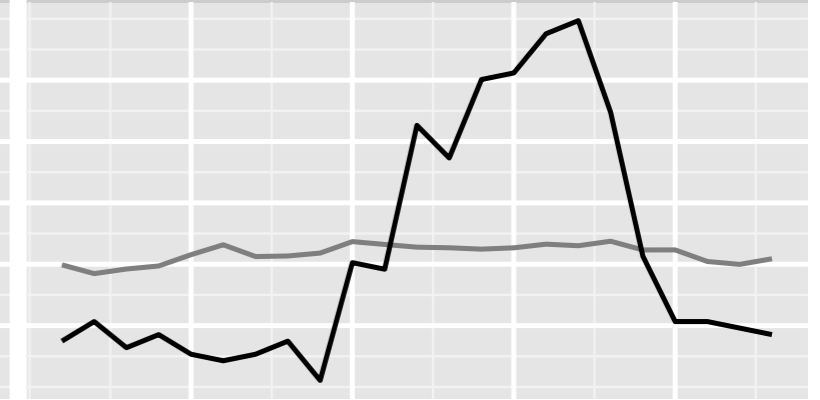
0.10  
0.08  
0.06  
0.04  
0.02



Assault (homicide) by sharp object

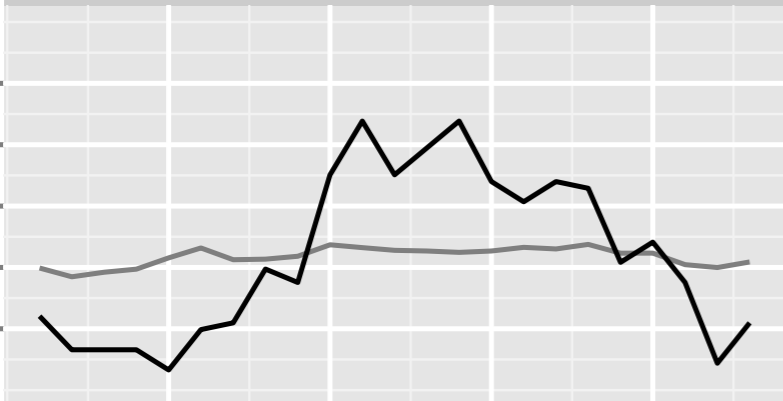


Drowning and submersion while in natural water

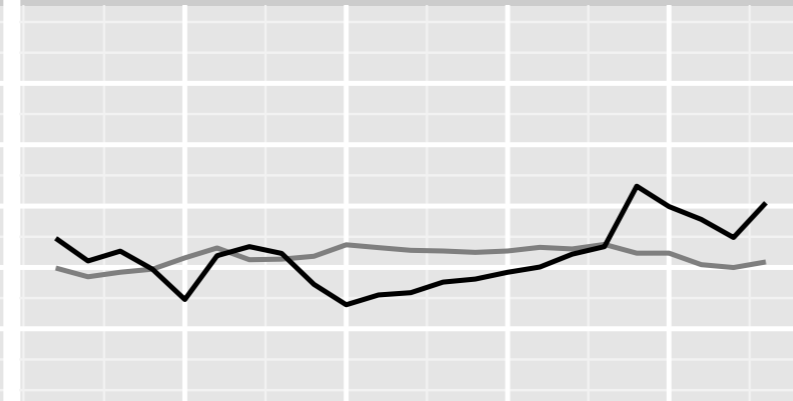


Exposure to unspecified electric current

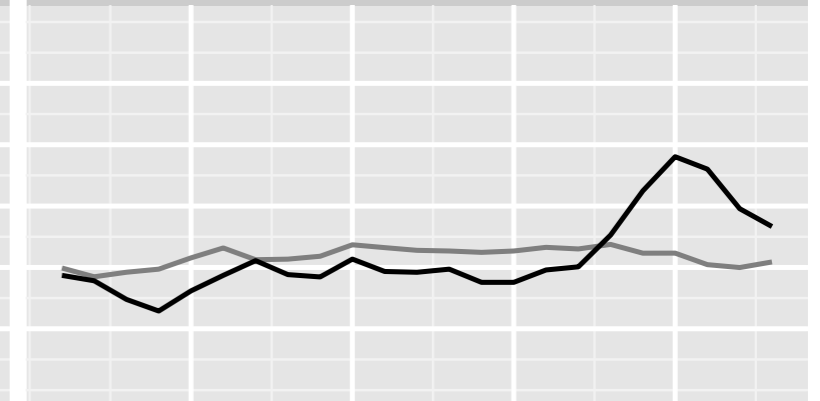
0.10  
0.08  
0.06  
0.04  
0.02



Motor- or nonmotor-vehicle accident, type of vehicle unspecified

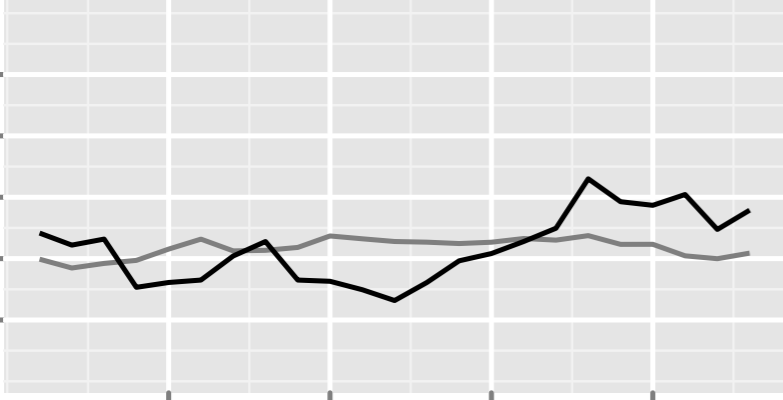


Pedestrian injured in other and unspecified transport accidents

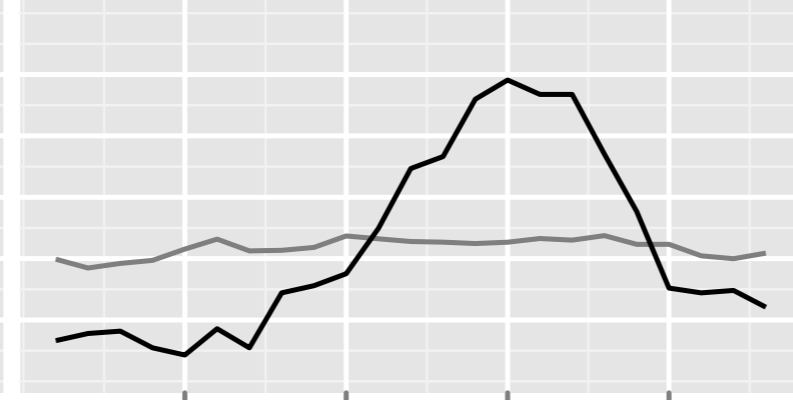


Traffic accident of specified type but victim's mode of transport unknown

0.10  
0.08  
0.06  
0.04  
0.02



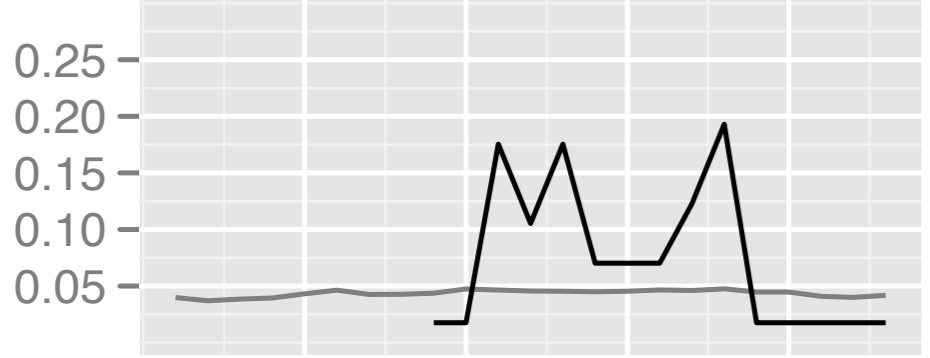
Unspecified drowning and submersion



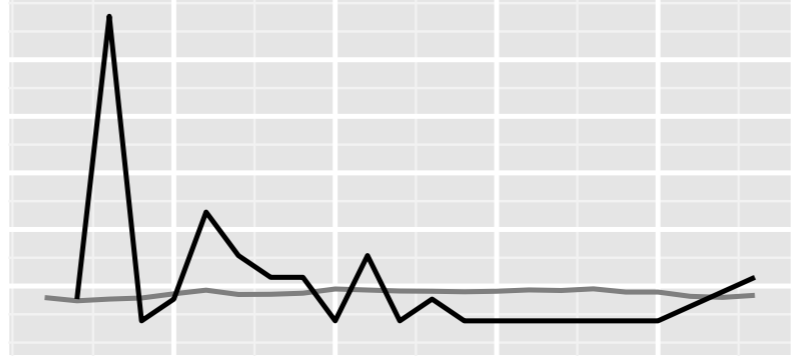
5 10 15 20

hod

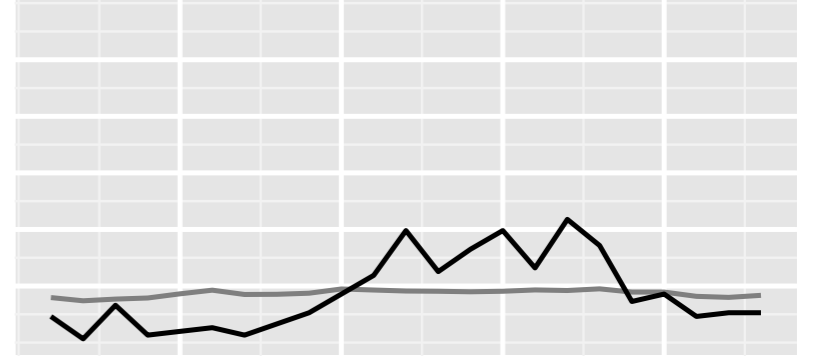
Accident to powered aircraft causing injury to occupant



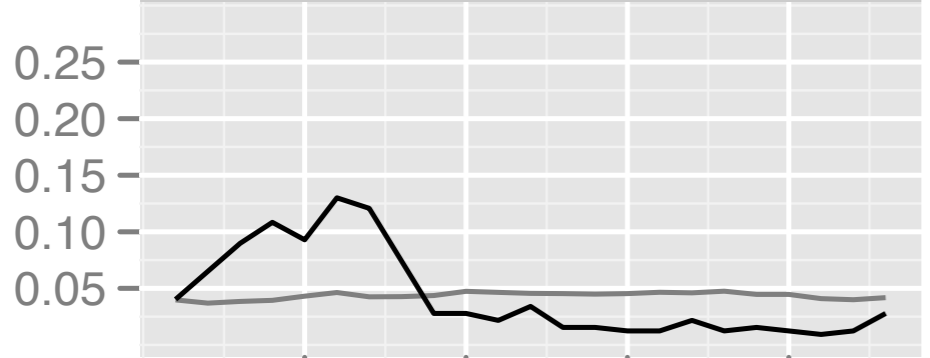
Bus occupant injured in other and unspecified transport accidents



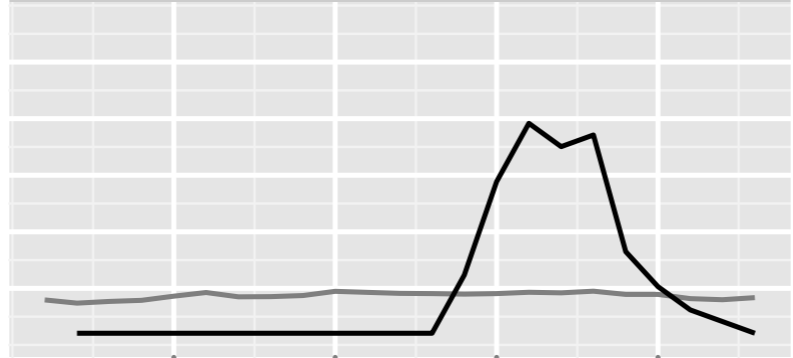
Other specified drowning and submersion



Sudden infant death syndrome



Victim of lightning



prop

hod

5 10 15 20



**Thanks!**





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.