Stat310

Continuous random variables

Hadley Wickham

1. Feedback
2. New conditions & definitions
3. Moments
4. Uniform distribution
5. Exponential and gamma

Feedback

Website

- HW 2 model answers now available.
- If you turn in a homework without a name on it, don't expect a grade.
- Practice problems for weeks 2 & 3.
- Notes up for all lectures.

```
----------------------------- attend class
```

----------------------------- attend class

- not getting totally discouraged

----------------------------- attend class

- not getting totally discouraged

--- not falling asleep

----------------------------- attend class

- not getting totally discouraged

--- not falling asleep

------------------------ paying attention/engaged

----------------------------- attend class

- not getting totally discouraged

--- not falling asleep

------------------------ paying attention/engaged

----------------- take notes

----------------------------- attend class

- not getting totally discouraged

--- not falling asleep

------------------------ paying attention/engaged

----------------- take notes

--------------- doing homework

```
----------------------------- attend class
- not getting totally discouraged
--- not falling asleep
------------------------ paying attention/engaged
----------------- take notes
--------------- doing homework
------- working through your turns
```

----------------------------- attend class

- not getting totally discouraged

--- not falling asleep

------------------------ paying attention/engaged

----------------- take notes

--------------- doing homework

------- working through your turns

--------------- starting hw early

----------------------------- attend class

- not getting totally discouraged

--- not falling asleep

------------------------ paying attention/engaged

----------------- take notes

--------------- doing homework

------- working through your turns

--------------- starting hw early

----- read textbook

----------------------------- attend class

- not getting totally discouraged

--- not falling asleep

------------------------ paying attention/engaged

----------------- take notes

--------------- doing homework

------- working through your turns

--------------- starting hw early

----- read textbook

------- collaborating/study group

----------------------------- attend class

- not getting totally discouraged

--- not falling asleep

------------------------ paying attention/engaged

----------------- take notes

--------------- doing homework

------- working through your turns

--------------- starting hw early

----- read textbook

------- collaborating/study group

------- reviewing notes

----------------------------- attend class

- not getting totally discouraged

--- not falling asleep

------------------------ paying attention/engaged

----------------- take notes

--------------- doing homework

------- working through your turns

--------------- starting hw early

----- read textbook

------- collaborating/study group

------- reviewing notes

--------------- understand better/faster/focus longer

----- take better notes

------- study more

--------------- understand better/faster/focus longer

----- take better notes

------- study more

---------------- start homework earlier

---------------- read text

---------- help sessions

---------------- understand better/faster/focus longer

----- take better notes

------- study more

---------------- start homework earlier

---------------- read text

---------- help sessions

------------------- review/update notes

--------------- understand better/faster/focus longer

----- take better notes

------- study more

---------------- start homework earlier

---------------- read text

---------- help sessions

------------------- review/update notes

--- get more sleep

-- ask questions in class

- work through previous homeworks
- don’t text during class

--------------- understand better/faster/focus longer

----- take better notes

------- study more

---------------- start homework earlier

---------------- read text

---------- help sessions

------------------- review/update notes

--- get more sleep

-- ask questions in class

- work through previous homeworks
- don’t text during class

Continuous random variables

Recall: have uncountably many outcomes

$$f(x)$$

For continuous x, f(x) is a probability **density** function.

Not a probability!

(may be greater than one)

$$P(a < X < b) = \int_a^b f(x)dx$$

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

What does P(X = a) = ?

$$P(a < X < b) = \int_a^b f(x)dx$$

To get probability, **always** need to integrate

What does P(X = a) = ?

PMF → PDF

Aside

Why do we need different methods for discrete and continuous variables?

Our definition of integration (Riemannian) isn't good enough, because it can't deal with discrete pdfs.

Can generalise to get Lebesgue integration. More general approach called measure theory.

CDF

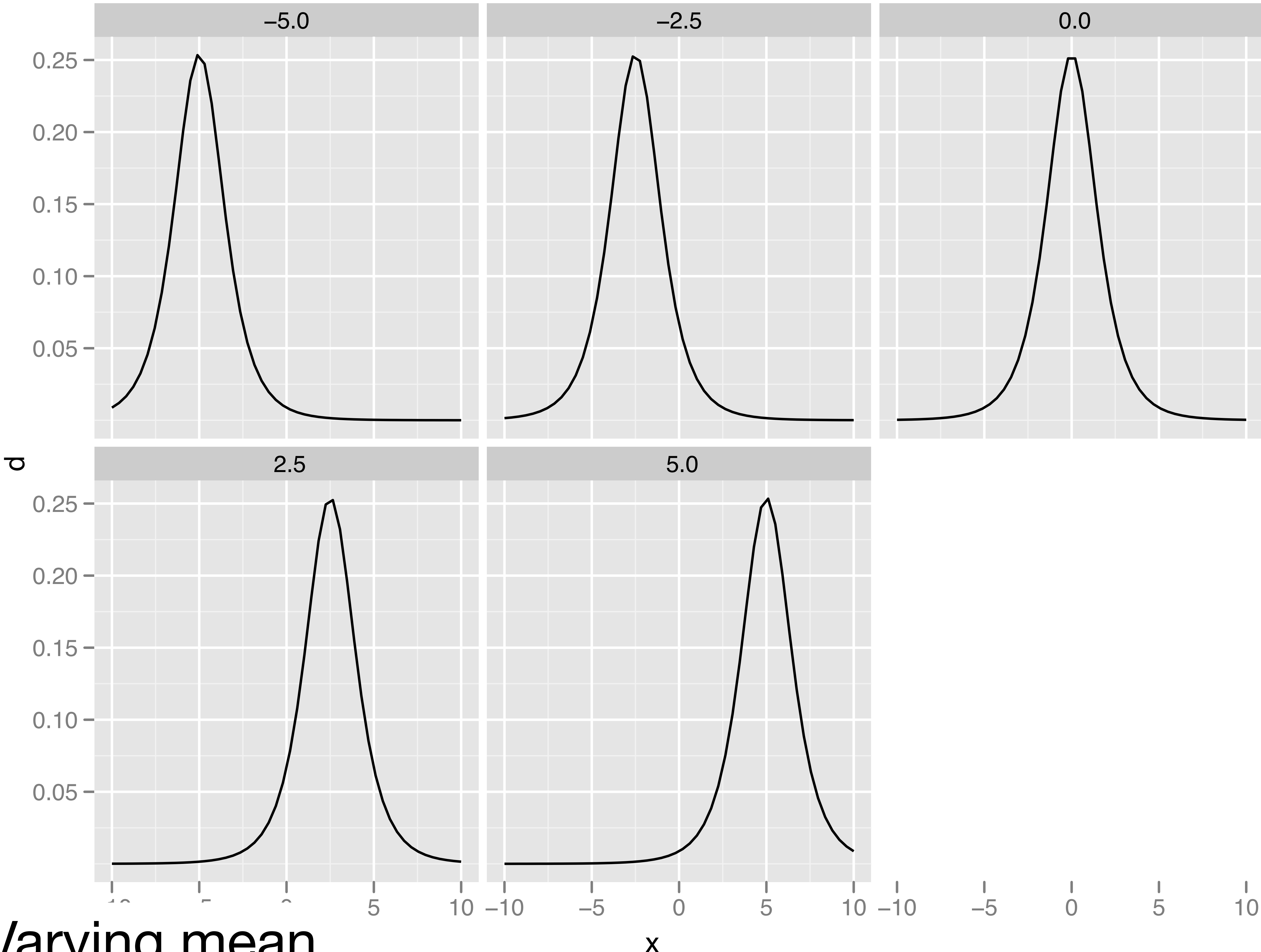
Cumulative distribution function

Useful because we're often interested in the probability of an interval

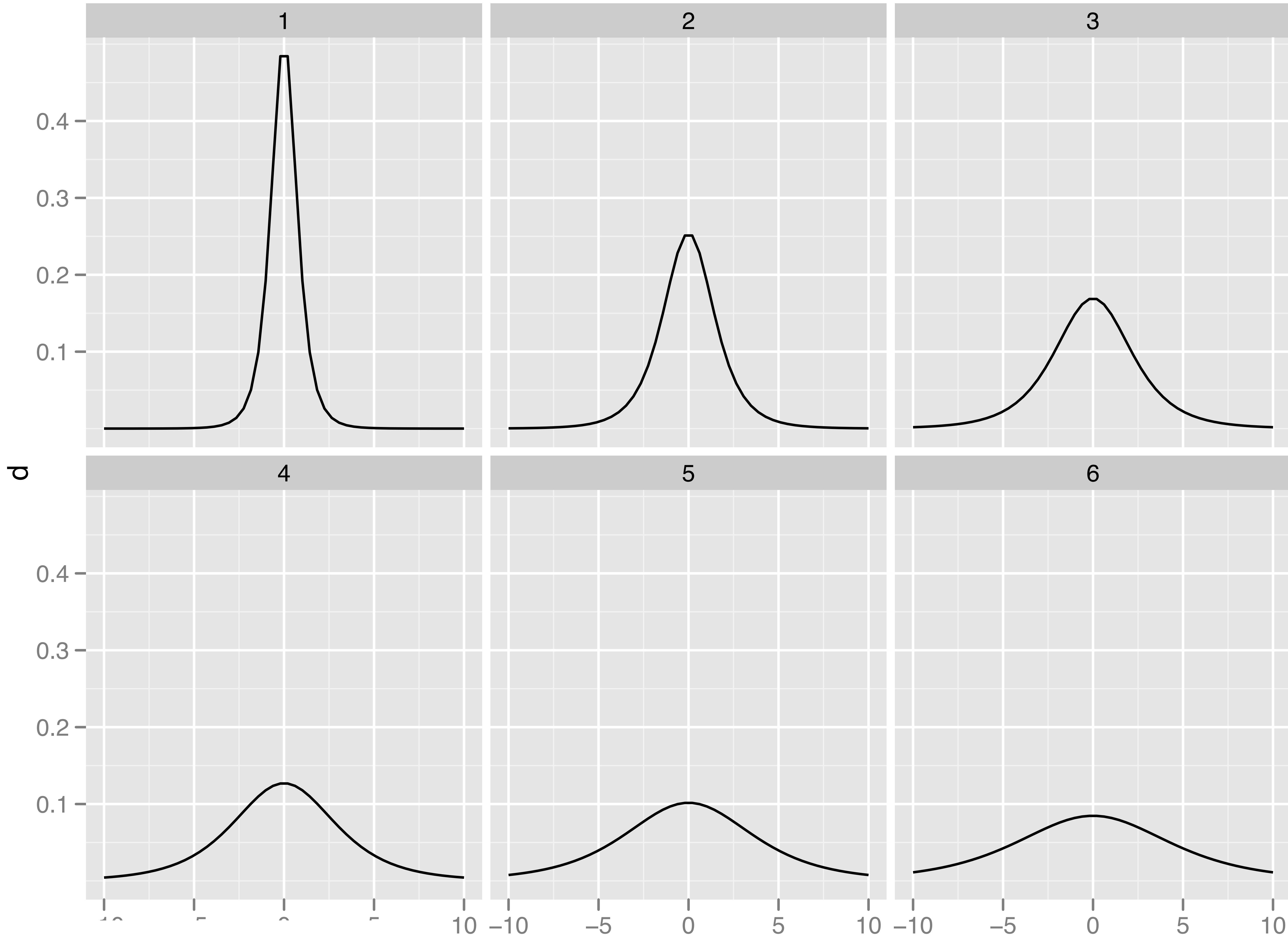
$$\begin{tikzcd} f(x) \arrow[d, bend right=60, "\text{Integrate from } -\infty \text{ to } x"'] \\ F(x) \arrow[u, bend right=60, "\text{Differentiate}"'] \end{tikzcd}$$

Moments

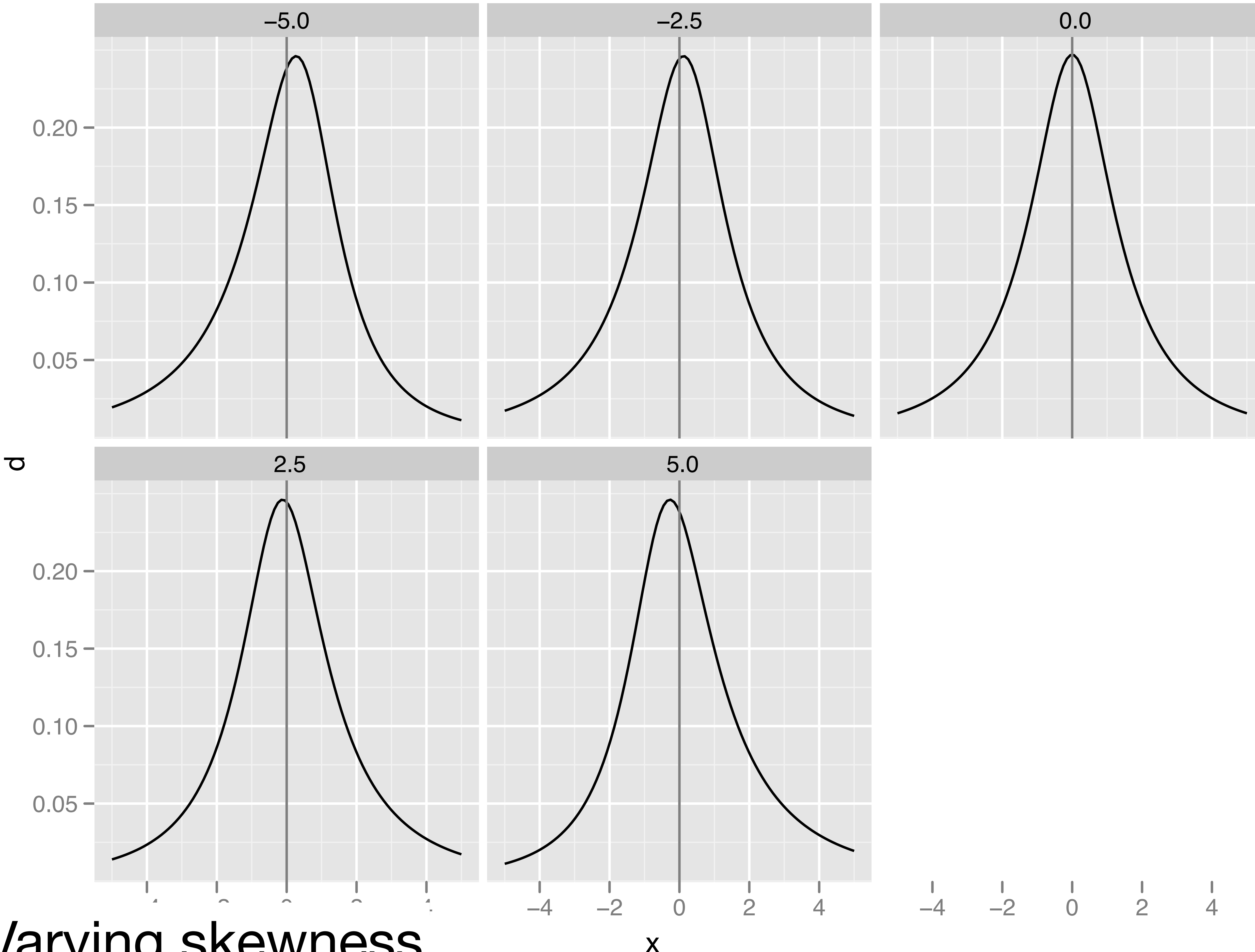
source("code/08-moments.r")



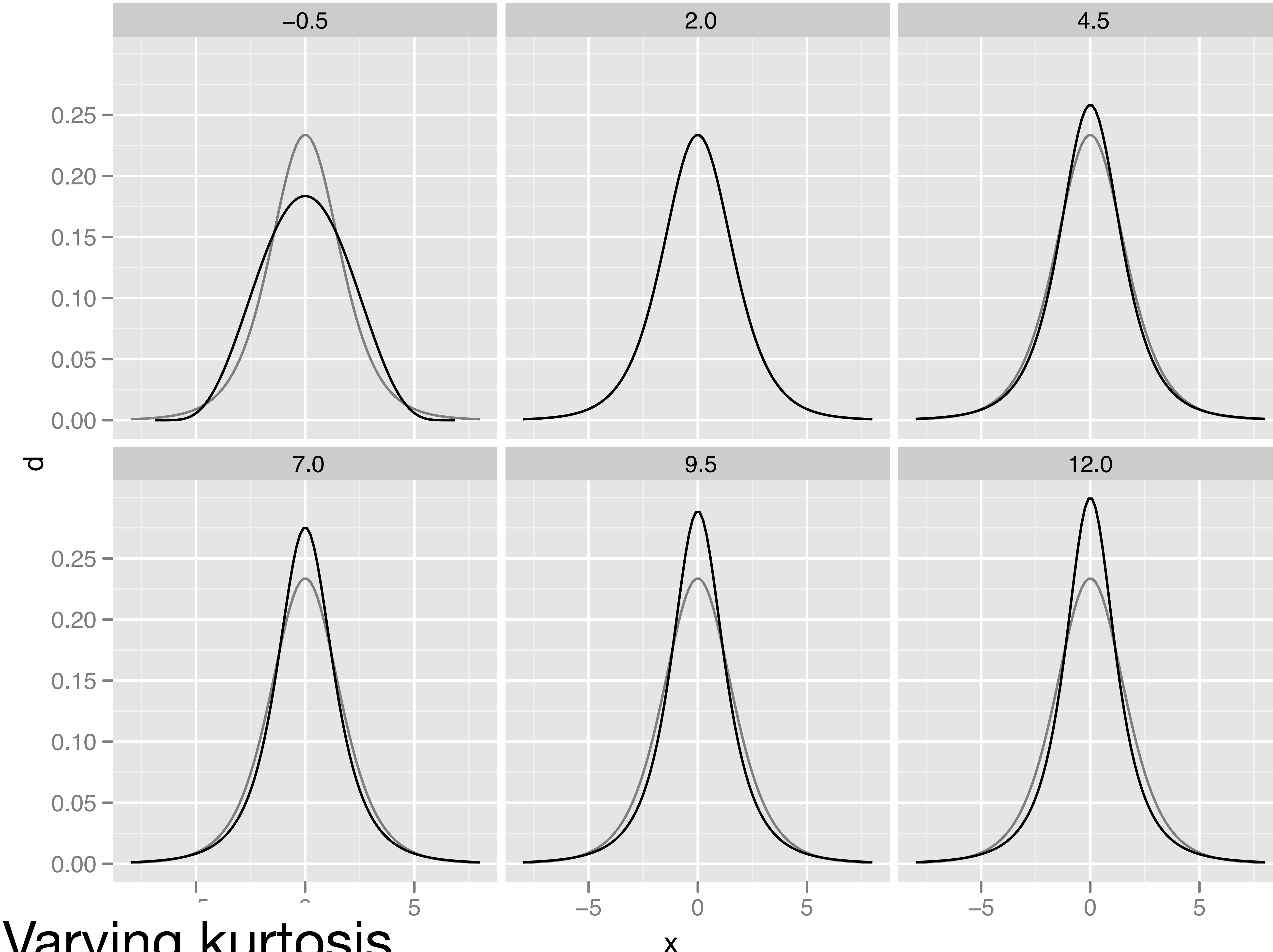
Varying mean



Varying variance



Varying skewness



Varying kurtosis

Uniform distribution

The uniform

Assigns probability uniformly in an interval [a, b] of the real line

X ~ Uniform(a, b)

What are F(x) and f(x) ?

$$\int_a^b f(x)dx = 1$$

Examples

Angle of spinner

Typically unrealistic, but a useful first start

(And can be transformed into any other distribution)

Description

- Assigns probability uniformly in an interval [a, b] of the real line
- X ~ Uniform(a, b)

$$
\begin{aligned}
F(x) &= 0 & x &< a \\
F(x) &= \tfrac{x-a}{b-a} & x &\in [a, b] \\
F(x) &= 1 & x &> b
\end{aligned}
$$

Intuition

$X \sim \text{Unif}(a, b)$

How do you expect the mean to be related to a and b? How do you expect the variance to be related to a and b?

(Think back to the discrete uniform)

$$M_X(t) = E[e^{Xt}] = \int_a^b e^{xt} \frac{1}{b-a} dx$$

```
integrate e^(x * t) 1/(b-a) dx from a to b
```

Mean & variance

Find mgf:

```
integrate e^(x * t) 1/(b-a) dx from a to b
```

Find 1st moment:

```
d/dt (E^(a t) - E^(b t))/(a t - b t) where t = 0
```

Find 2nd moment:

```
d^2/dt^2 (E^(a t) - E^(b t))/(a t - b t) where t = 0
```

Find var:

```
simplify 1/3 (a b+a ^2+b^2)  - ((a + b)/2) ^ 2
```

$$E(X) = \frac{a+b}{2}$$

$$Var(X) = \frac{(b-a)^2}{12}$$

Exponential and gamma distributions

Conditions

Beware variations!

Let X ~ Poisson(λ).
Average waiting time: $1 / \lambda$

Let Y be the time you wait for the next event to occur. Then
Y ~ **Exponential**($\theta = 1 / \lambda$)

Let Z be the time you wait for α events to occur. Then Z ~ **Gamma**(α, $\theta = 1 / \lambda$)

But generally, the exponential and gamma are a useful approximation for many waiting times - it’s a good first try.

http://www.causeweb.org/repository/statjava/Distributions.html#GAMMA

	Until one	Multiple
Count	geometric	negative binomial
Waiting time	exponential	gamma

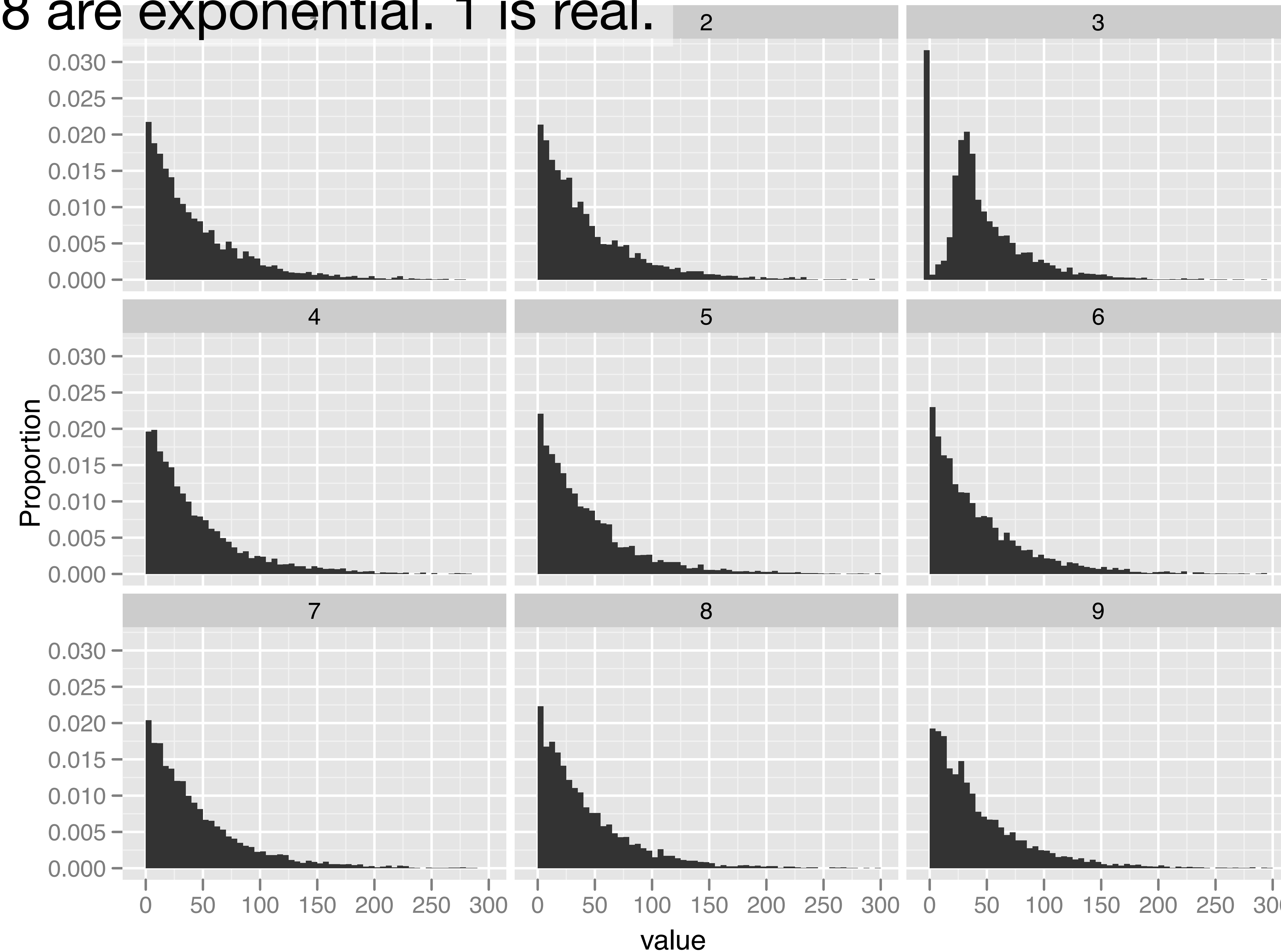
LA Lakers

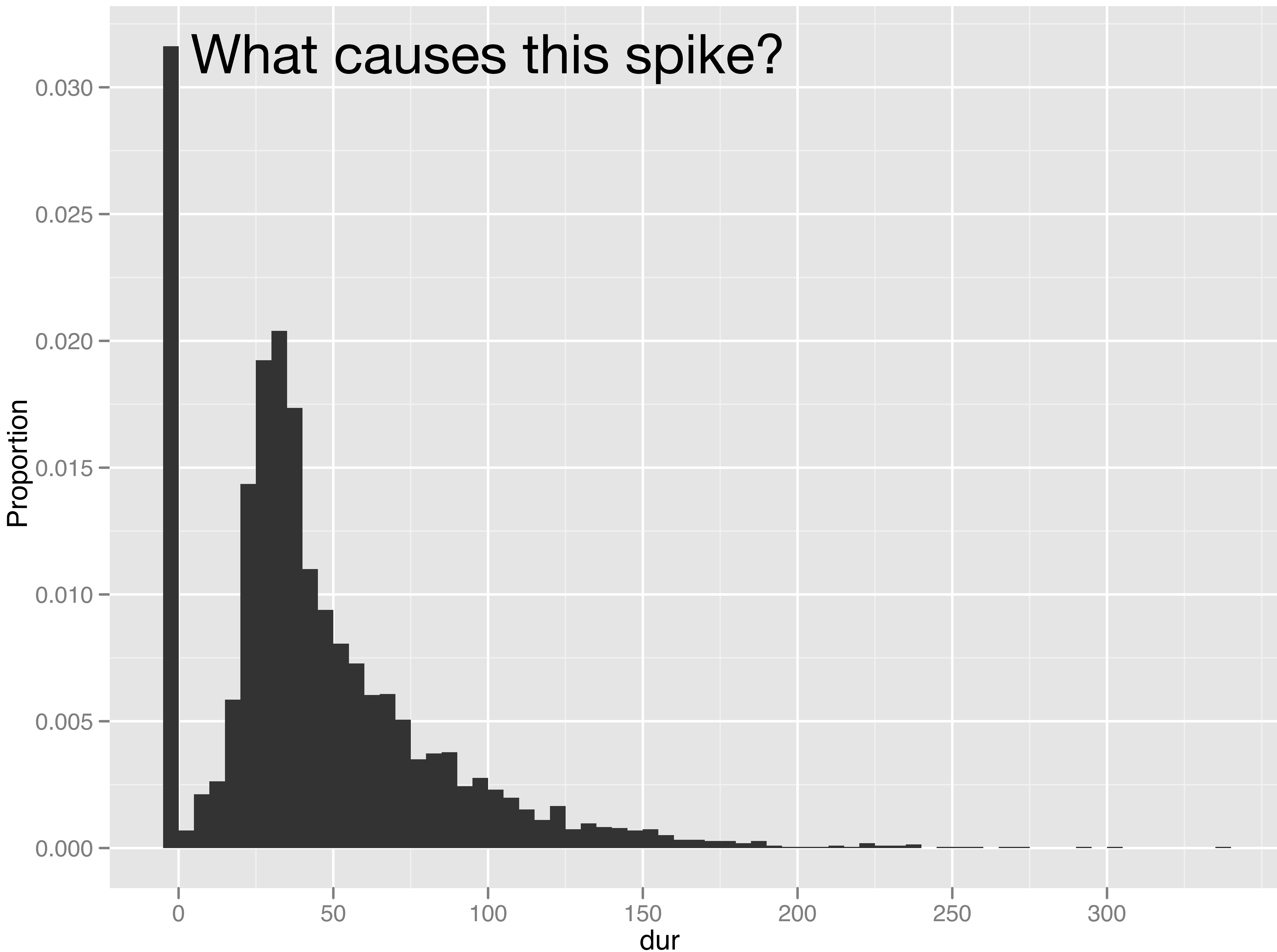
Last time we said the distribution of scores looked pretty close to Poisson. Does the distribution of times between scores look exponential?

(Downloaded play-by-play data for all NBA games in 08/09 season, extracted all point scoring plays by Lakers, computed times between them - 4660 in total)

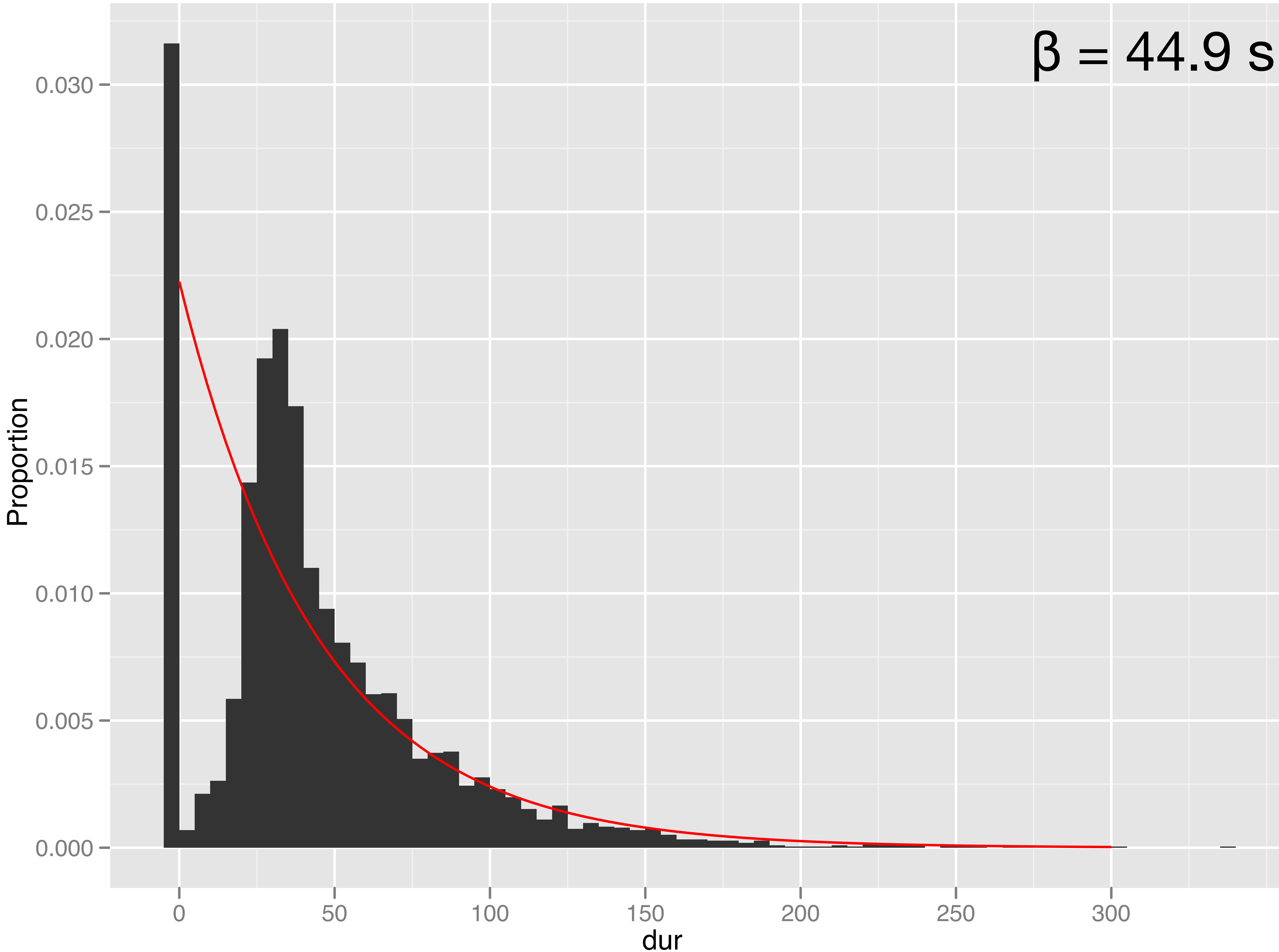
If you want to learn how to do this yourself, take stat405

8 are exponential. 1 is real.





What causes this spike?
Proportion
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0
50
100
150
200
250
300
dur


β = 44.9 s
Proportion
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
50
100
150
200
250
300
dur

Why?

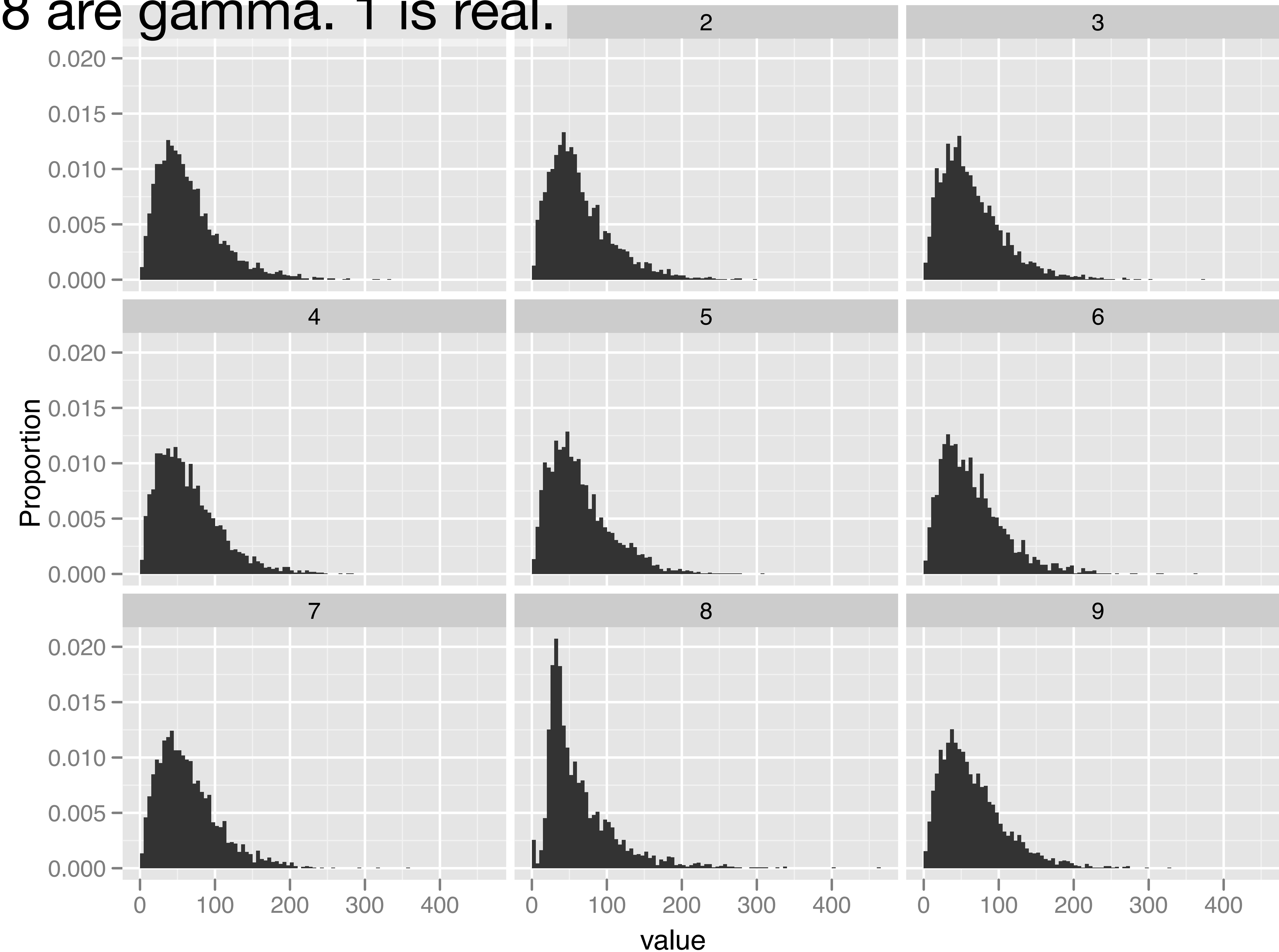
Why is it easy to see that the waiting times are not exponential, when we couldn't see that the scores were not Poisson?

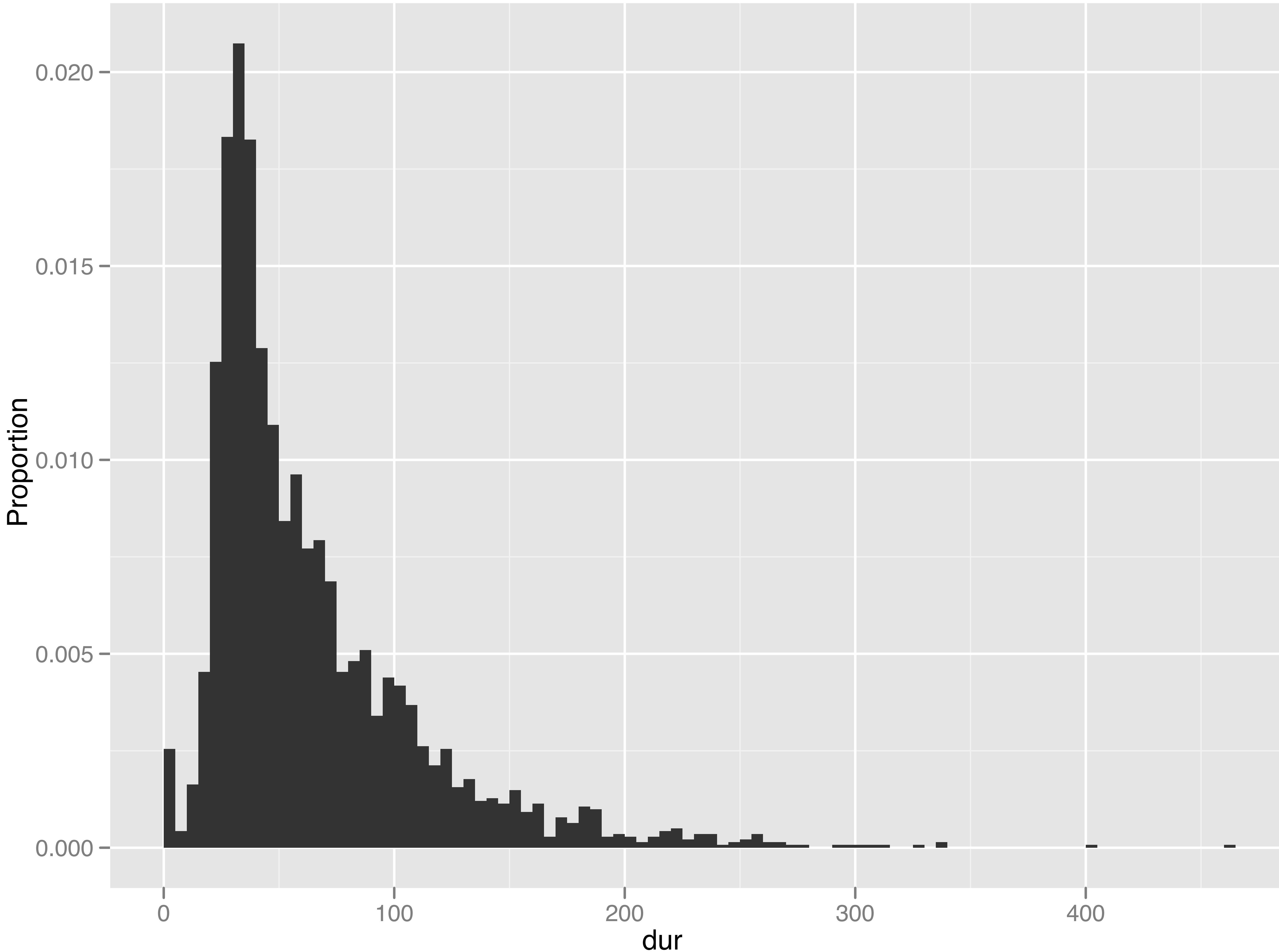
Next

Remove free throws.

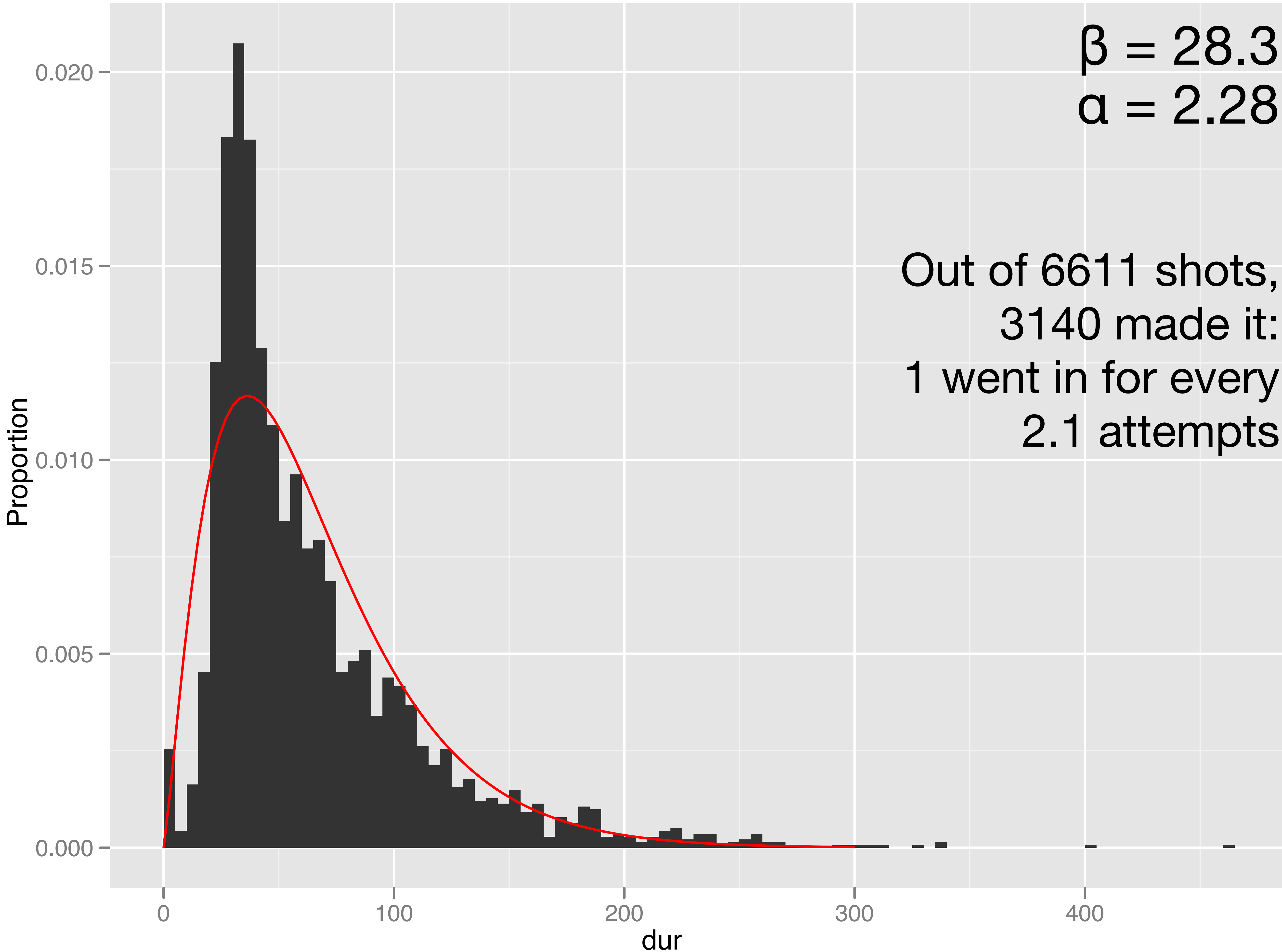
Try using a gamma instead (it's a bit more flexible because it has two parameters, but it maintains the basic idea)

8 are gamma. 1 is real.





0.020
0.015
0.010
0.005
0.000
Proportion
0
100
200
300
400
dur


β = 28.3
α = 2.28
Out of 6611 shots,
3140 made it:
1 went in for every
2.1 attempts
Proportion
0.020
0.015
0.010
0.005
0.000
0
100
200
300
400
dur

pdf $$f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}$$

mgf $$M(t) = \frac{1}{(1-\theta t)^\alpha}$$

mean $$\alpha\theta$$

variance $$\alpha\theta^2$$

What is gamma?

```
integrate theta^(-alpha) x^(alpha -
1) exp(-x / theta) from 0 to infinity
```

Also: a continuous analog of the factorial: $\Gamma(n) = (n-1)!$

http://en.wikipedia.org/wiki/Gamma_function

Finding the mean

Find mgf: integrate 1 / (gamma(alpha) theta^alpha) x^(alpha - 1) e^(-x / theta) e^(x*t) from x = 0 to infinity

Find mean: d/dt theta^(-alpha) (1/theta - t)^(-alpha) where t = 0

Rewrite: simplify alpha (1/theta)^(-alpha) theta^(1-alpha)

Your turn

Assume the distribution of offensive points by the LA lakers is Poisson(109.6). A game is 48 minutes long.

Let W be the time you wait between the Lakers scoring a point. What is the distribution of W?

How long do you expect to wait?