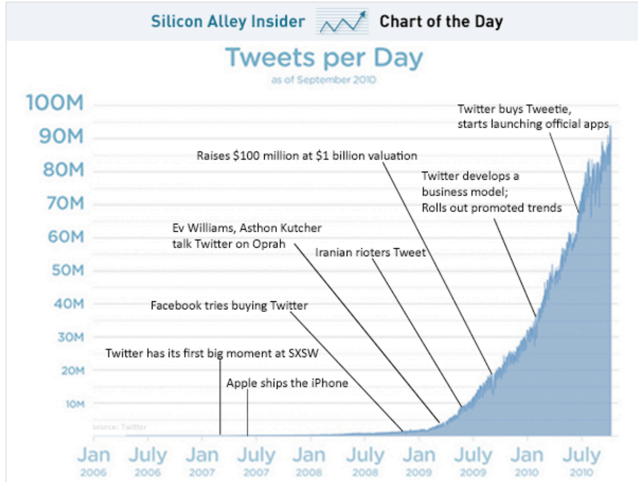
Activity #12: Modeling Assignments
Topics: Creating, choosing, evaluating mathematical models

I (@Thiessen) joined Twitter on May 24, 2008 as user #14,891,817. This means I joined right before @Kaspinuotis (a Lithuanian blogger) and immediately after @JeremySkelly (a UX Developer from Cleveland). The first Twitter user, as far as I can tell, was Jack Dorsey (the creator of Twitter) with an ID number of 12.

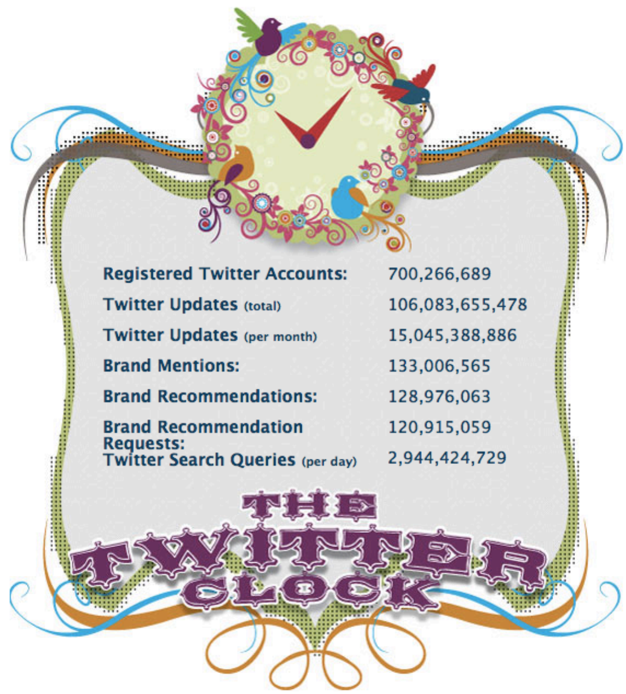
Realizing that ID numbers must be directly related to the number of users, I became interested in the number of Twitter users over time. I searched online and found the graph displayed to the right, but I could not find a table showing the number of Twitter *users*.

Finally, I stumbled across *The Twitter Clock*, a site that shows real-time Twitter usage data. The screenshots below display the data from this page as of 9:42 and 9:45 AM on November 5, 2012.



9:42 AM

9:45 AM



I captured this data several times from 9:42 AM to 4:46 PM and pasted it in the following table:

Date	Time	Minutes past 9:42	Accounts	Tweets	Brand Mentions
11/5/12	9:42 AM	0	700,265,069	106,083,438,218	133,005,949
11/5/12	9:43 AM	1	700,265,609	106,083,510,638	133,006,154
11/5/12	9:44 AM	2	700,266,149	106,083,583,058	133,006,360
11/5/12	9:45 AM	3	700,266,689	106,083,655,478	133,006,565
11/5/12	9:52 AM	10	700,270,469	106,084,162,418	133,008,000
11/5/12	10:02 AM	20	700,275,869	106,084,886,618	133,010,052
11/5/12	10:42 AM	60	???	???	???
11/5/12	4:46 PM	424	???	???	???

1) Your goal is to fill-in-the-blanks on the previous page to predict future numbers of accounts, tweets, and brand mentions. To do this, we're going to fit linear and exponential functions to the data.

Before we begin, let's simplify our data a bit. We don't expect the number of Twitter accounts to go up by 1 million in an hour, so let's focus on hundreds of thousands. Likewise, let's ignore some of the larger place values in the number of tweets and brand mentions:

Date	Time	minutes	Accounts 700,000,000 +	Tweets 106,000,000,000 +	Brand Mentions 133,000,000 +
11/5/12	9:42 AM	0	265,069	83,438,218	5,949
11/5/12	9:43 AM	1	265,609	83,510,638	6,154
11/5/12	9:44 AM	2	266,149	83,583,058	6,360
11/5/12	9:52 AM	10	270,469	84,162,418	8,000
11/5/12	10:02 AM	20	275,869	84,886,618	10,052

Let's first fit a linear model to this data to predict the number of accounts at 10:42 AM. To do this, let's arbitrarily pick two points and rewrite them as ordered pairs:

9:42 AM: Accounts = 265,069 -----> (0, 265069)
10:02 AM: Accounts = 275,869 -----> (20, 275869)

From these two points, calculate the slope and y-intercept to find the equation of the linear function. Write the slope and y-intercept into the blanks below.

Accounts = ________________ (minutes past 9:42) + ________________ (+ 700,000,000)
Y = m X + b

Interpret the slope you just calculated. What does it represent?

____

Use the same data (9:42 and 10:02) to calculate tweets and brand mentions, each as a function of time. Write the equations in the blanks below:

Tweets = ________________ (minutes past 9:42) + ________________ (+ 106,000,000,000)

Brand Mentions = ________________ (minutes past 9:42) + ________________ (+ 133,000,000)

2) Evaluate the 3 linear functions you wrote above to predict the number of accounts, tweets, and brand mentions on Twitter at 10:42 AM and 4:46 PM. Write your answers in the following blanks:

Date	Time	Minutes	Accounts 700,000,000 +	Tweets 106,000,000,000 +	Brand Mentions 133,000,000 +
11/5/12	10:42 AM	60	________	________	________
11/5/12	4:46 PM	424	________	________	________

3) Let's now fit an exponential function to the data. We know exponential functions model situations where Y has a constant percentage growth, so we need to estimate the growth rate of accounts, tweets, and brand mentions.

To do this, let's look at the data at 9:42 and 9:43. Calculate the percent increase in accounts, tweets, and brand mentions over this one-minute time period. Write these percentage increases in the following blanks:

Minutes	Accounts 700,000,000 +	% Change	Tweets 106,000,000,000 +	% Change	Brand Mentions 133,000,000 +	% Change
0	265,069		83,438,218		5,949	
1	265,609	____________	83,510,638	____________	6,154	____________

Using these growth rates, write out the exponential functions:

Accounts = ________________ (____________________) (minutes past 9:42) (+ 700,000,000)

Y = a (1+r) ^t

Tweets = ________________ (____________________) (minutes past 9:42) (+ 106,000,000,000)

Brand Mentions = ________________ (____________________) (minutes past 9:42) (+ 133,000,000)

4) Evaluate the 3 exponential functions you wrote above to predict the number of accounts, tweets, and brand mentions on Twitter at 10:42 AM and 4:46 PM. Write your answers in the following blanks:

Date	Time	Minutes	Accounts 700,000,000 +	Tweets 106,000,000,000 +	Brand Mentions 133,000,000 +
11/5/12	10:42 AM	60	____________	____________	____________
11/5/12	4:46 PM	424	____________	____________	____________

5) Of the two sets of predictions you've made (linear and exponential), which do you think will be most accurate? Why?

____

____

____

6) The following table displays the actual data reported by *The Twitter Clock* at 10:42 AM and 4:46 PM that day.

Date	Time	Minutes	Accounts	Tweets	Brand Mentions
11/5/12	10:42 AM	60	700,297,469	106,087,783,418	133,018,257
11/5/12	4:46 PM	424	700,494,038	106,114,145,505	133,092,943

Compared to the actual data, which of your functions (linear or exponential) appears to best describe the number of accounts, tweets, and brand mentions as a function of time?

Circle One: Linear Function Exponential Function

Explain why you think the function you circled above best describes the data.

____

____

____

7) Based on your linear and exponential models, at what time did I take the screenshot to the right?

According to the linear model, this screenshot was taken

________________ minutes after 9:42 AM.

According to the exponential model, this screenshot was taken

________________ minutes after 9:42 AM.

Sources:

Find ID number from a user name: http://www.idfromuser.com

Find user name from ID: https://twitter.com/account/redirect_by_id?id=14891817
(The number at the end, 14891817, is the ID number you want to look-up)

Twitter Clock: http://tweetreports.com/brand-monitoring/twitter-clock-twitter-brand-monitoring/

Twitter growth chart: http://wewillraakyou.com/twitter-account-growth-dynamic-graph/

8) In 2009, Usain Bolt broke the world record in the 100 meters event. During that race, he reached a top speed of 27.79 miles per hour.

Why can't I run that fast? Some might say it's because I'm out-of-shape and only run when something's chasing me. I disagree. I think it's because I'm simply not tall enough. Usain Bolt is 4 inches taller than me, so that must be why he runs faster.

Biologists have long known that there is an important mathematical relationship between the running speed of animals and their overall body length. The following table gives the lengths of various organisms, in centimeters, and their top running speed in centimeters per second:

Animal	Length (cm)	Speed (cm/sec)
Clover mite	0.08	0.85
Ant	0.42	6.50
Deer mouse	9.00	250.00
Zebra-tail lizard	15.00	720.00
Chipmunk	16.00	480.00
Iguana	24.00	730.00
Gray Squirrel	25.00	760.00
Red Fox	60.00	2000.00
Cheetah	120.00	2900.00

Source: McMahon, T.A. & Bonner, J.T. (1983). On Size and Life, Scientific American Library.

Do the following:

a) Using length as the independent variable and speed as the dependent variable, sketch a scatterplot.

b) Using a calculator, find the **linear** function that best fits the data. Write out the equation for that linear function, report the R^2 value, and briefly discuss how well the function models the data.

c) Use your linear function to predict the speed of each animal in the table. How accurate is the best prediction? How accurate is the worst prediction?

d) Use your calculator to find an **exponential** function that best fits the data. Write out the equation for that exponential function, report the R^2 value, and briefly discuss how well the function models the data.

e) Use your calculator to find a **quadratic** function that best fits the data. Write out the equation for that exponential function, report the R^2 value, and briefly discuss how well the function models the data.

f) Use your calculator to find a **power** function that best fits the data (STAT -> CALC -> PWRREG). Write out the equation for that exponential function, report the R^2 value, and briefly discuss how well the function models the data.

g) Decide which of the three functions best describes the relationship between the length of an animal and its top speed. Discuss the criteria you used to make this judgment.

h) A full-size house cat is about 28 cm long. Using the function you chose in part (g), estimate the top speed of a house cat. A Google search informs me the top speed of a house cat is 1,332 cm/sec. How accurate was your estimate?

i) A full-size lion can run at a top speed of 2,235 cm/sec (50 mph). Using the function you chose in part (g), estimate the length of a full-size lion. A Google search informs that a lion can be between 200 and 300 cm long. How accurate was your estimate?

j) Usain Bolt is 196 cm tall. Using the function you chose in part (g), estimate his top speed.

k) Usain Bolt's top speed is 1,241 cm/sec. Using the function you chose in part (g), estimate his height. How accurate were your estimates?