Performance Based Learning and Assessment Task

Modeling Statistical Data

I. ASSESSMENT TASK OVERVIEW & PURPOSE:
The task is to provide students with a chance to investigate real life data to analyze a real world situation and use their knowledge of functions to make connections and predictions.

II. UNIT AUTHOR:
Amber Bebout, Radford High School, Radford City Schools

III. COURSE:
Algebra, Functions, and Data Analysis

IV. CONTENT STRAND:
Data Analysis

V. OBJECTIVES:
The learner will be able to:

- Use collected data from an online resource to create tables and graphs
- Use the created graphs to predict the equation that fits the data most accurately through finding a curve of best fit
- Explain and justify his/her reasoning behind the chosen curve and equation and use it to predict data for a future time frame

VI. REFERENCE/RESOURCE MATERIALS:
Students will need a copy of the activity prompt. Each student will need access to a computer, unless they work in groups, then one computer per group should be adequate. Graphing calculators or an equivalent application is necessary.

VII. PRIMARY ASSESSMENT STRATEGIES:
The task includes an assessment component that the student may use as a checklist for self-assessment and the teacher may use as a grading rubric. The assessment component is made up of two parts: a mathematical component and a communication component.

VIII. EVALUATION CRITERIA:

- Mathematical Component will be evaluated using Making Sense of Math Rubric
- Communication Component will be evaluated using Data Presentation Rubric
- Benchmark of Exemplary Work will guide teachers for intended baseline of students' work

IX. INSTRUCTIONAL TIME:
This activity should take four 50-minute class periods.

Modeling Statistical Data

Strand
Data Analysis

Mathematical Objective(s)
- Students will use their knowledge of functions to analyze data and make predictions.
- Students will correctly use supportive data to validate their predictions and conclusions

Related SOL
AFDA.3 The student will collect data and generate an equation for the curve (linear, quadratic, exponential, and logarithmic) of best fit to model real-world problems or applications. Students will use the best fit equation to interpolate function values, make decisions, and justify conclusions with algebraic and/or graphical models.

NCTM Standards
The student will:
- Apply and adapt a variety of appropriate strategies to solve problems
- Communicate their mathematical thinking coherently and clearly to peers, teachers, and others
- Compute fluently and make reasonable estimates
- Develop and evaluate inferences and predictions that are based on data
- Understand relations and functions and select, convert flexibly among, and use various representations for them
- Use the language of mathematics to express mathematical ideas precisely.
- Use representations to model and interpret physical, social, and mathematical phenomena

Materials/Resources
- Graphing Calculators, Excel, GeoGebra, or similar program capable of graphing and calculating equations of best fit
- Computers (lab or cart; for individual students or one per group of students)
- *Optional: Pre-Activity Worksheet on Fitting Curves and Analyzing Data
- This worksheet can be edited to fit any district/city the teacher desires. The worksheet based on Virginia, District 4 and Radford City as that is the creator's current location.
- Making Sense of Math & Presentation Rubric
- Activity Prompt

Assumption of Prior Knowledge
- The student must be comfortable using an appropriate program to find the best equation related to a set of data and using that to find the equation of best fit. Example programs: GeoGebra, Excel, or Desmos.
- The student must know the basic parent graphs and how to write equations for linear, quadratic, exponential, and logarithmic functions.
- The student should know how to use the models they will create to make a prediction based on facts.

Further Extension for Classroom Discussion
The teacher may use the following information to extend the activity prompt with further class discussion
- The teenage pregnancy rates are solely based on the number of pregnant teens within the given age range(s) and do not account for married teens.
- The non-marital birth rates are based on marriage status upon delivery, not whether the father is known/unknown.

Introduction: Setting Up the Mathematical Task

In this activity, you will investigate the relationship between collected data and how to use it to make predictions for future time frames. You will collect data from the Virginia Department of Health website based on the given prompt and with this data, you will create tables and graphs in order to explore curve of best fit and to predict future events. You will use graphing calculators or a similar computer program that graphs to assist in the problem solving process, but you should not fully rely on the calculator, as you will need to interact with the material to thoughtfully complete all components of the activity. You will work individually on this project until all students are finished. At that point, each student will present their conclusions in the manner in which they have chosen to fulfill the presentation requirements.

Student Exploration

Whole Class Discussion (estimated time: 10-15 minutes)

- The teacher will introduce the activity by facilitating a discussion on how data is collected and for what purpose organizations may want specific data collected.
- The teacher will then show students the website from which they will be collecting data. A discussion can follow on the different categories that data is given for on the website.
- The teacher will then provide students with the activity prompt.

Individual Work (estimated time: 2-3 class periods)

- Students will work individually on the project with assistance from the teacher when necessary

Student Presentation (estimated time: 1 class period)

- Students will use this time to turn in their final products with the self-assessment included for teacher grading.

Student/Teacher Actions:

- Students should take into careful consideration the data they are collecting and the best way to analyze the data.
- Students should carefully examine the activity prompt and the provided rubrics to ensure they understand the expectations of the activity.
- Students may not completely grasp how to begin the activity and how best to display the data; if so, the teacher should provide guidance through mentioning use of tables and graphs to get started.
- The teacher should circulate and ensure students understand what their goal is for the collected data. The teacher can ask questions such as: “What would a scatter plot of that data look like?” “Is there a better way to display the data you have collected?” “What is the data telling you?”
- The teacher may need to redirect students on their conclusions if students are not making conclusions/predictions based on facts.
- Students will present their data and prediction(s) they are claiming in an organized manner.

Monitoring Student Responses

- Students should be supportive and respectful of all students during the activity. Providing help to struggling students should be encouraged, but should not simply be for the supplying of answers.
- If students need a modification to the activity, providing them with the tables of information is an option.
- Students will be encouraged and expected to use mathematical vocabulary during the activity and in their activity presentation.
- The activity will close with each student presenting their findings and drawing conclusions based on fact.

*Optional

Pre-Activity Worksheet on Fitting Curves and Analyzing Data

Collecting Data on DIVORCE RATES

Pre Assessment Question: What are your current thoughts on divorce? Do you think there is cause for concern pertaining to this topic? Explain your answer with 2-3 sentences.

In the following table, record the appropriate data for the total number of divorces for the State of Virginia, District 4, and Radford City.

How to get to the data:

1. Open the following webpage: http://www.vdh.virginia.gov/healthstats/stats.htm
2. Scroll down until you see "Recorded Divorces by City/County of Occurrence."
3. In the drop down menu to the right, choose the year and record the appropriate data in the chart.

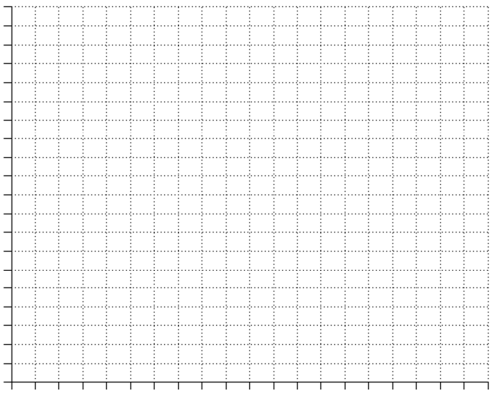
Data Collection Table 1

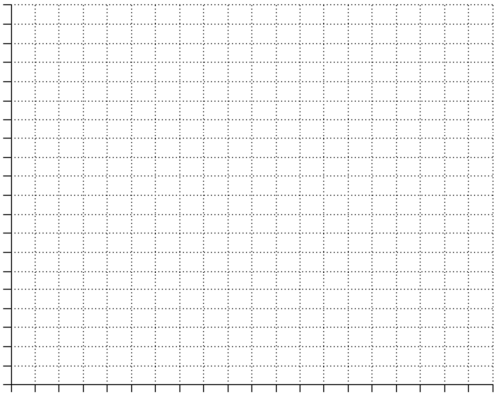
Year	State of Virginia	District 4	Radford City
2000			
2001			
2002			
2003			
2004			
2005			
2006			
2007			
2008			
2009			
2010			
2011			
2012			
2013			

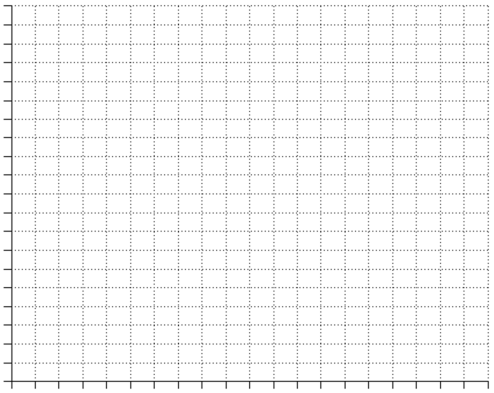
1. Based on the data you collected in Table 1, what can you conclude about the relationship between the year and the number of divorces in the State of VA? District 4? Radford City?

2. If you had to make a prediction for each location for the year 2020, what do you think the number of divorces will be then?

3. For the State of Virginia, District 4, and Radford City, create a graph that makes the number of divorces in a given year a function of the year. Clearly label your graph and explain why you choose the unit of measurement you used.







4. Compare the three graphs you just created based on Table 1. What similarities do you see? What differences do you see?

5. Draw a curve of best fit for each graph. Then write an equation that correctly fits each set of data. Clearly identify your variable choices.

6. Explain and justify your reasoning for choosing the curve of best fit you used for each graph.

7. Is there another curve of best fit that could be applied to the State of Virginia? District 4? Radford City? Explain.

8. Using the equations found in #5, now make an educated prediction for the years 2015, 2020, 2025.

Data Collection Table 2

Year	State of Virginia	District 4	Radford City
2015			
2020			
2025			

9. How do your data in #8 for the year 2020 compare to your original prediction in #2? Explain why it is close, exact, or far off by taking into account your original reasoning behind your answer to #2.

10. Based on your equation, will there be a time that there are no more divorce occurrences in the State of Virginia? District 4? Radford? If yes, what year(s)?

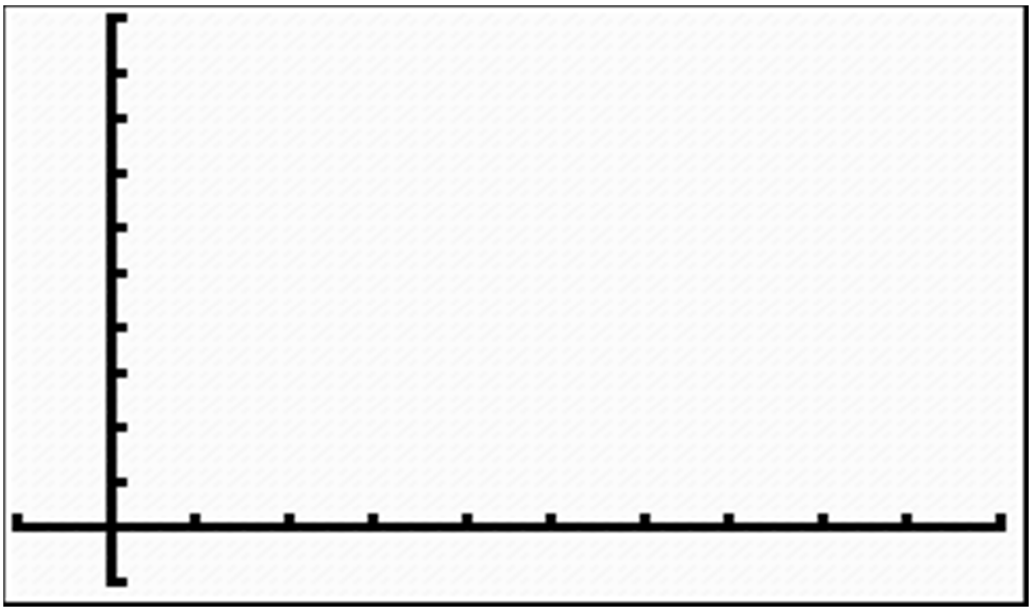
11. What are some factors that may affect the number of divorces? Explain your reasoning for each factor.

12. You'll notice that the document where you originally found the data for Table 1 also has columns for the number of divorces separated by the number of children involved. Use the data to fill out the following table, Table 3, for the State of Virginia ONLY.

Data Collection Table 3

Year	Total Number of Divorces	0 Children	1 Child	2 Children	3 Children	4 or more Children	Unknown
2000							
2001							
2002							
2003							
2004							
2005							
2006							
2007							
2008							
2009							
2010							
2011							
2012							
2013							

13. Create a graphical representation to compare any variation in the number of divorces among the 7 different categories for the number of children involved. Clearly label the graph so that a clear comparison can be seen.



14. Based on your graph in #13, are there any generalizations you can make about how the number of divorces relates to the number of children involved? Provide a thoughtful explanation for the reasoning behind your each generalization you make.

Post Assessment Question: How has working through the data on divorce occurrences affected your thoughts on divorce? Has it changed your mind on how frequently they occur? Explain your thoughts and reactions after working through this activity with 3-5 sentences.

Answer Key

Pre-Activity Worksheet on Fitting Curves and Analyzing Data

Collecting Data on the NUMBER OF DIVORCES

Pre Assessment Question: What are your current thoughts on divorce? Do you think there is cause for concern pertaining to this topic? Explain your answer with 2-3 sentences.

[Answers will vary]

In the following table, record the appropriate data for the total number of divorces for the State of Virginia, District 4, and Radford City.

How to get to the data:

1. Open the following webpage: http://www.vdh.virginia.gov/healthstats/stats.htm
2. Scroll down until you see "Recorded Divorces by City/County of Occurrence."
3. In the drop down menu to the right, choose the year and record the appropriate data in the chart.

Data Collection Table 1

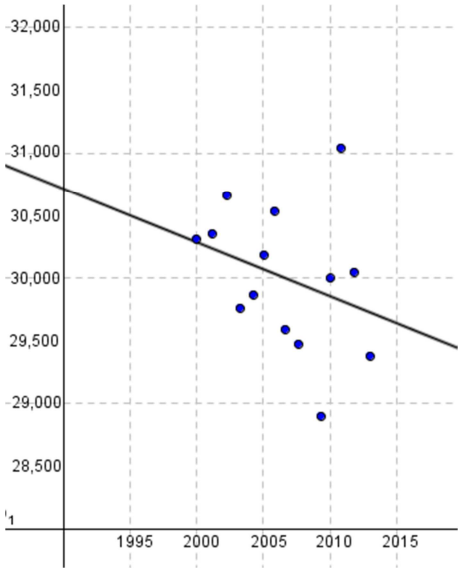
Year	State of Virginia	District 4	Radford City
2000	30,360	661	74
2001	30,448	707	47
2002	30,833	740	66
2003	29,743	623	59
2004	29,814	621	42
2005	30,186	629	31
2006	30,606	576	27
2007	29,519	634	34
2008	29,460	584	39
2009	28,979	576	36
2010	30,011	619	27
2011	31,068	651	43
2012	30,048	622	32
2013	29,465	506	31

1. Based on the data you collected in Table 1, what can you conclude about the relationship between the year and the number of divorces in the State of VA? District 4? Radford City?
   The relationship between the year and the number of divorces for the State of VA seems to remain fairly constant with only a slight change considering the number of divorces. For District 4, the number of divorces also seems to be consistent. For Radford City, the number of divorces had declined and appears to be staying around the thirty to forty range.

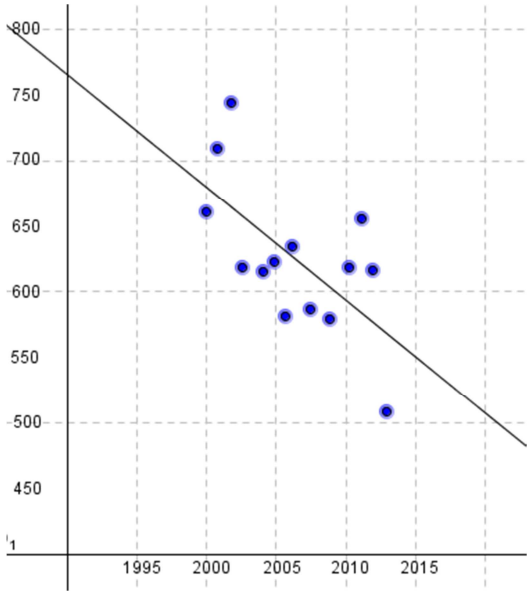
2. If you had to make a prediction for each location for the year 2020, what do you think the number of divorces will be then?
   [Answers will vary]
   State of Virginia: 31,020
   District 4: 587
   Radford City: 36

3. For the State of Virginia, District 4, and Radford City, create a graph that makes the number of divorces in a given year a function of the year. Clearly label your graph and explain why you choose the unit of measurement you used.

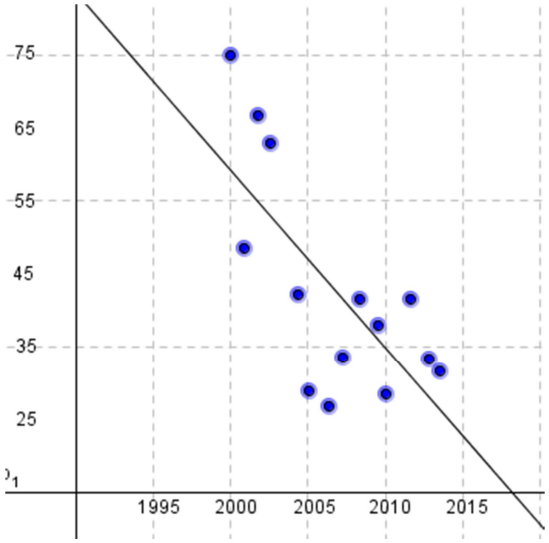
State of Virginia



District 4



Radford City



4. Compare the three graphs you just created based on Table 1. What similarities do you see? What differences do you see?
   Each graph has points that are clustered together with a few outliers. The outliers for District 4 and Radford City seem to be in about the same time frame, whereas the State of Virginia's outliers come from a few years later.

5. Draw a curve of best fit for each graph. Then write an equation that correctly fits each set of data. Clearly identify your variable choices.
   Let y be the number of divorces and x be the year in which the data was collected:
   State of Virginia: $y = -42.82x + 115,960$
   District 4: $y = -8.97x + 18,621.69$
   Radford City: $y = -2.54x + 5131$

6. Explain and justify your reasoning for choosing the curve of best fit you used for each graph.
   I chose linear for each of the three graphs because there wasn't another definite curve of best fit that fit the data closer. If there were years' worth of data below year 2000 and above 2013, a more definite trend would more likely be seen and possibly a better curve of best fit. The linear curve was also chosen because the r^2 value was closet to one.

7. Is there another curve of best fit that could be applied to the State of Virginia? District 4? Radford City? Explain.
   No, linear seems to be the best. Each of the graphs has outliers, but it definitely appears, for the years we are taking into account, that the relationship is linear.

8. Using the equations found in #5, now make an educated prediction for the years 2015, 2020, 2025.

Data Collection Table 2

Year	State of Virginia	District 4	Radford City
2015	29,678	548	13
2020	29,464	503	1
2025	29,250	458	-13

9. How do your data in #8 for the year 2020 compare to your original prediction in #2? Explain why it is close, exact, or far off by taking into account your original reasoning behind your answer to #2.
   [Answers may vary] My prediction for the State of Virginia is far off considering the values originally found for the years 2000-2013. I overestimated thinking the trend would swing back in favor of divorces based on the higher values I saw in 2001, 2006, 2011. My prediction for District 4 was fairly close, but I under estimated the decline District 4 would potentially see in the years following 2013. This makes me think that maybe the population in District 4 declined by 2020; hence the number of divorces would also decline. My prediction for Radford City is way off. I based my original prediction on thinking Radford City would still have a decent population in 2020. It would seem that either Radford City residents' acceptance of divorce, number of divorce lawyers, or population has decreased since 2013.

10. Based on your equation, will there be a time that there are no more divorce occurrences in the State of Virginia? District 4? Radford? If yes, what year(s)?
    Yes. The State of Virginia will see a divorce rate of zero in 2709, District 4 will see it in year 2076, and Radford City will see it in year 2021. Models do fail though and can exist only within certain parameters. Therefore, even though these years would be the projected time for the number of divorces will be zero, actually having the number of divorce be zero when the human race still exists is not reasonable.

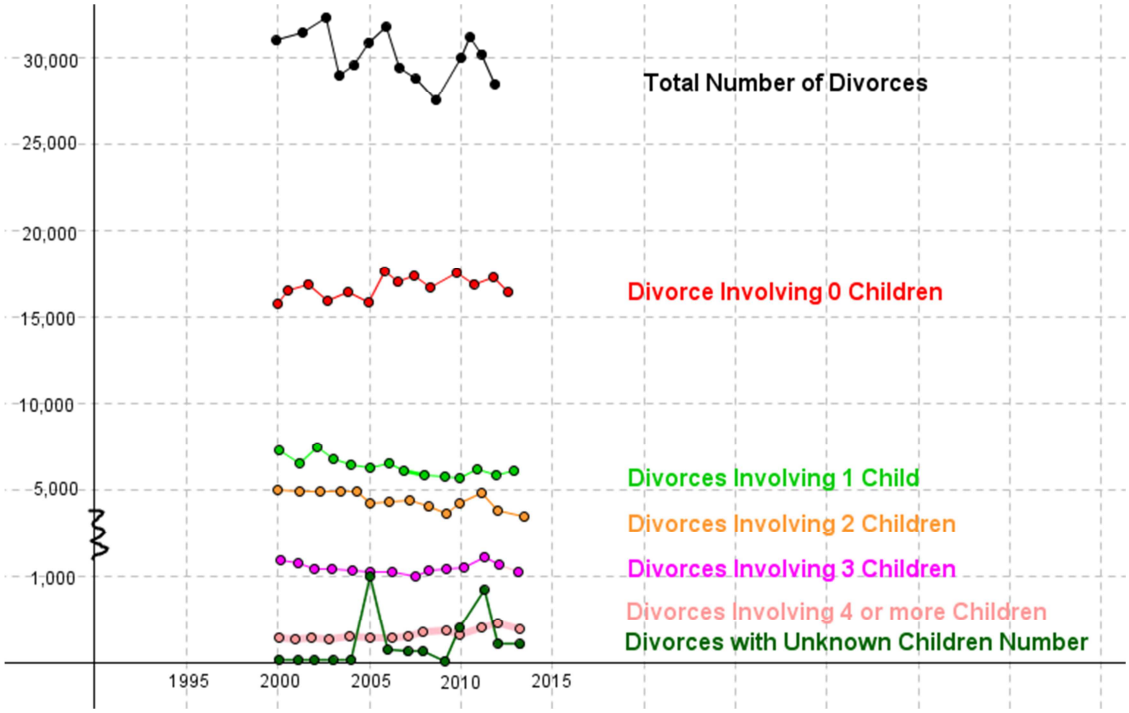
11. What are some factors that may affect the number of divorces? Explain your reasoning for each factor.
[Answers will vary] Social Acceptance: Throughout the years, divorce has turned from being something that was not socially acceptable to being more acceptable. It hasn't come fully into acceptance, but the initial shock of someone coming from a split home has decreased and, in some areas of the United States, coming from a split home seems more like a norm. Divorce rates will probably plateau as society spends less time focusing on them and it just becomes a norm.
Number of Initial Marriages: The history of marriage has changed in the minds of men and women around the world in the last century. Many more couples are deciding to forego marriage and cohabitate as it is becoming more acceptable in society. Fewer marriages would result in fewer possible divorces.
Rise in Stricter Religious Beliefs: Some religions do not accept divorce for their members and so a rise in the population belonging to these religions could have an effort on how many divorces occur.

12. You'll notice that the document where you originally found the data for Table 1 also has columns for the number of divorces separated by the number of children involved. Use the data to fill out the following table, Table 3, for the State of Virginia ONLY.

Data Collection Table 3

Year	Total Number of Divorces	0 Children	1 Child	2 Children	3 Children	4 or more Children	Unknown
2000	30,360	16,111	7,400	5,191	1,307	290	61
2001	30,448	16,575	7,291	5,060	1,230	271	21
2002	30,833	16,709	7,419	5,147	1,271	276	11
2003	29,743	16,066	7,162	4,995	1,256	254	10
2004	29,814	16,433	6,986	4,908	1,181	287	19
2005	30,186	16,136	6,917	4,651	1,167	268	1,047
2006	30,606	17,193	6,955	4,788	1,255	278	137
2007	29,519	16,665	6,590	4,721	1,141	280	122
2008	29,460	16,788	6,419	4,633	1,238	273	109
2009	28,979	16,555	6,347	4,537	1,170	299	62
2010	30,011	17,043	6,399	4,758	1,247	272	292
2011	31,068	16,761	6,748	5,005	1,353	328	873
2012	30,048	16,841	6.628	4,762	1,287	338	192
2013	29,465	16,756	6,454	4,478	1,223	308	246

13. Create a graphical representation to compare any variation in the number of divorces among the 7 different categories for the number of children involved. Clearly label the graph so that a clear comparison can be seen.



14. Based on your graph in #13, are there any generalizations you can make about how the number of divorces relates to the number of children involved? Provide a thoughtful explanation for the reasoning behind your each generalization you make.
[Answers will vary] Each category for the number of divorces based on the number of children seems to remain fairly constant through the years 2000 to 2013. The graph clearly shows that more divorces happen among couples that have no children. This I believe is supported by the thinking behind the importance of family. Once children are involved, you'll notice that the number of divorces occurring drastically decreases with a very low number occurring for families with four or more children, comparatively speaking. Based on the number of divorces with an unknown number of children it can be assumed that, during the years 2005 and 2010-2013, there was a higher rate of divorces that did not report the number of children involved, as these years have elevated results compared to other years. The year 2005 contains the largest number at 1,047. This lack of reporting may be due to the divorces ending peacefully in terms of visitation rights, child support, etc.

Post Assessment Question: How has working through the data on divorce occurrences affected your thoughts on divorce? Has it changed your mind on how frequently they occur? Explain your thoughts and reactions after working through this activity with 3-5 sentences.
[Answers will vary]

Activity Prompt
Making Conclusions from a Data List

You are given the task of collecting data from the following website and making a prediction for the year 2016. The data you collect will be based on either the Number of Teen Pregnancies or the Number of Non-Marital Birth. You must collect data for the State of Virginia, a district of your choice, and a city of your choice within the aforementioned district. You should make visuals that relate to the data and the equation for the curve of best fit, draw conclusions about the data sets, and make the required prediction based on facts. Upon completion of your exploration, you will put together a written two page paper (visuals are welcome, but written content must be two pages, double spaced), a PowerPoint intended to be self-guiding, or verbal presentation through PowerPoint or another medium.

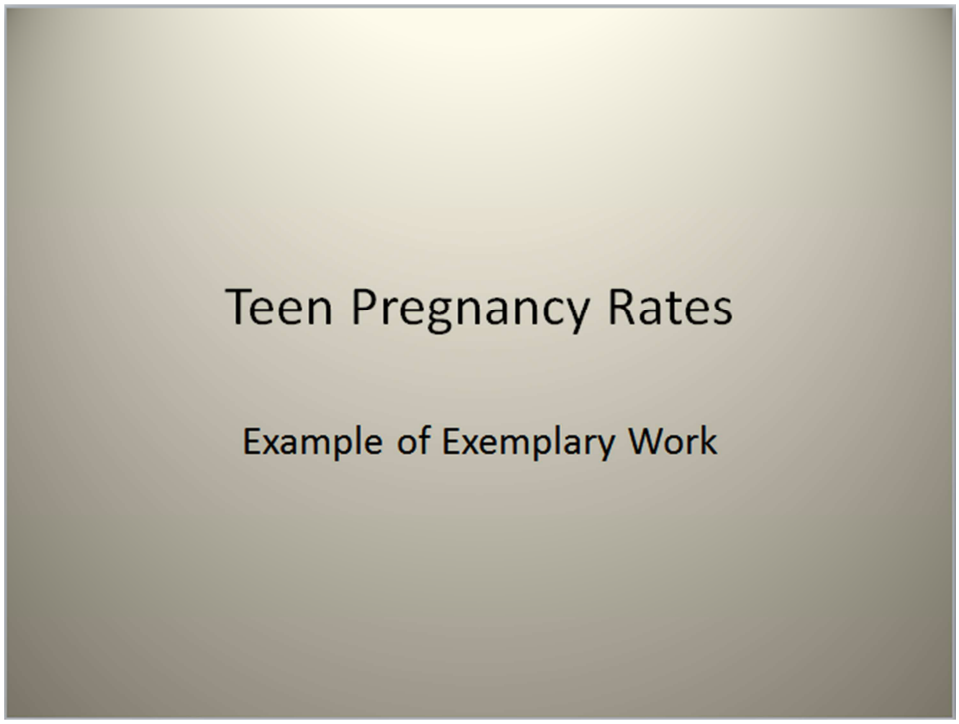
Website: http://www.vdh.virginia.gov/healthstats/stats.htm

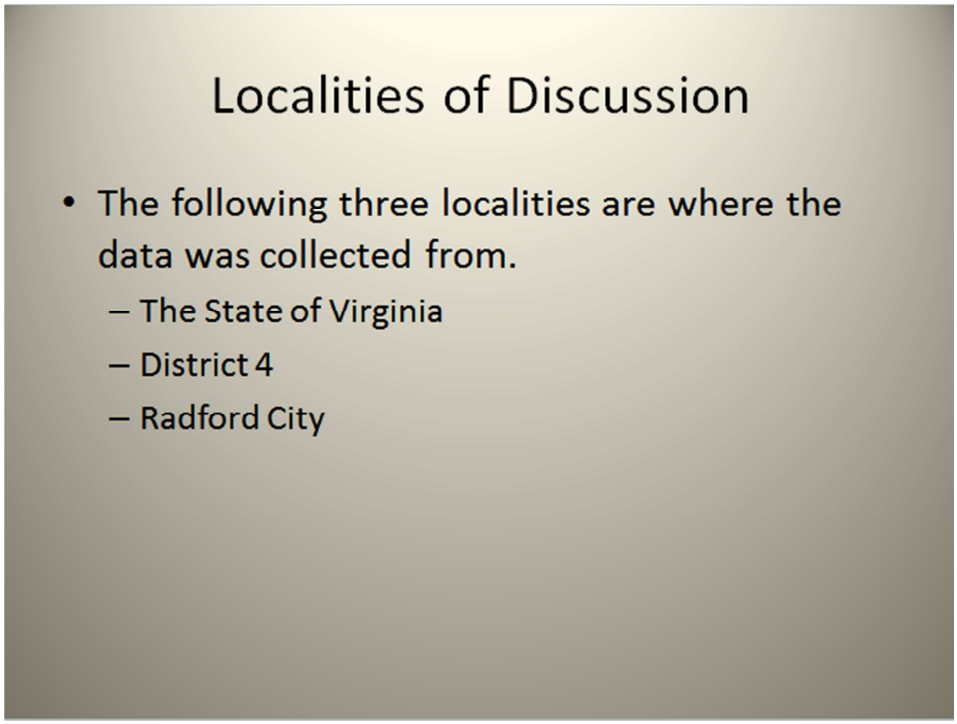
Activity Assessment List and Benchmarks

Making Sense of Math & Presentation Rubric

	0	1	2	Student Assessment	Teacher Assessment
Making Sense of Math Portion					
Student uses appropriate method for analyzing data	The student did not know how to analyze the data	The student analyzed the data but did not use an proper method	The student analyzed the data through a proper method	/2	/2
Student clearly collects and labels data	The students did not collect/label the data	The student collected/ labeled the data but in an unorganized manner	The student collected/labeled the data in an organized, easy to understand manner	/2	/2
Student correctly finds relationship between variables	The student did not find a relationship between the variables	The student found a relationship between the variables but it was inaccurate	The student found an accurate relationship between the variables	/2	/2
Student writes equations based on the data	The student did not write equations and/or the equations are inaccurate	The student wrote equations, but they are not all accurate	The student wrote accurate equations where needed	/2	/2
Student's calculations are accurate	The student has more than 3 errors in their calculations	The student has 1-2 errors in their calculations	The student has no errors in their calculations	/2	/2
Data Presentation Portion					
Student has chosen a medium for presenting from those provided	The student has chosen to use a medium not appropriate for the presentation of data	- - -	The student uses a medium of presentation permitted	/2	/2
Student has included visuals that support their presentation's intention(s)	The student uses no visuals in their presentation	The student uses visuals, but they do not fully support their presentation's intention(s)	The student uses visuals that support and enrich their presentation	/2	/2
Student's graphs are easy to interpret and accurate	The student's graphs lack labeling and/or are hard to decipher relationship between variables	The student made graphs, but omitted important information such as axis labels, measurement labels, curve of best fit	The student made graphs that are clearly labeled and is an accurate representation of the collected data	/2	/2
Student uses the found relationship to draw accurate conclusions	The student did not draw conclusions	The student drew inaccurate conclusions due to misunderstanding the data	The student drew accurate conclusions supported by the data	/2	/2
Student makes prediction that is reasonable and based on facts	The student did not make a prediction	The student made a prediction but it is not based on facts or supported by the data found	The student made a predication that is clearly supported by the data	/2	/2
			Total	/20	/20

Making Conclusions from a Data List
Sample of Exemplary Work
Medium: Self-Guided PowerPoint



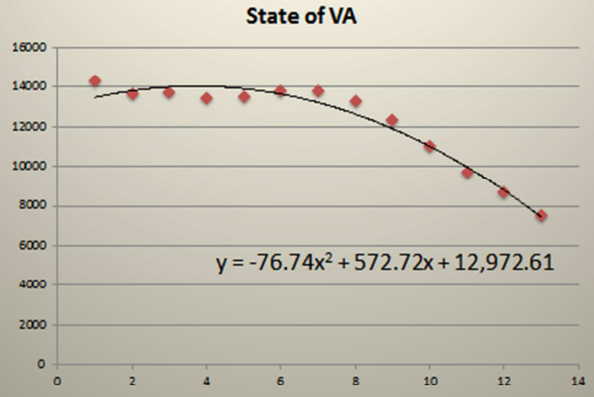


Collected Data

Year	State of VA	District 4	Radford City
2001	14218	275	42
2002	13586	246	37
2003	13665	233	37
2004	13343	238	33
2005	13416	238	26
2006	13704	202	27
2007	13766	231	25
2008	13233	282	37
2009	12283	248	25
2010	10970	218	27
2011	9630	174	30
2012	8651	212	33
2013	7447	164	21

Based on the collected data, one can see that, for all three localities, the number of teen pregnancies seems to be decreasing.

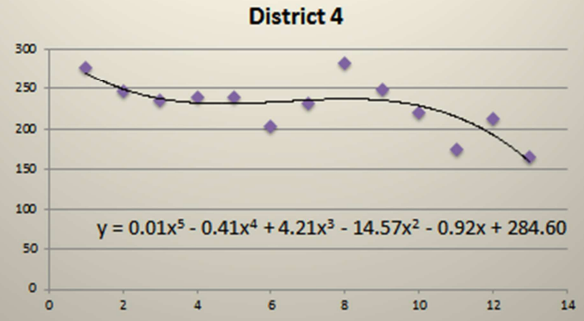
Individual Scatterplots & Equation of Best Fit



For the State of Virginia, a quadratic polynomial was used to model the data as it clearly follows the data and had the highest correlation coefficient.

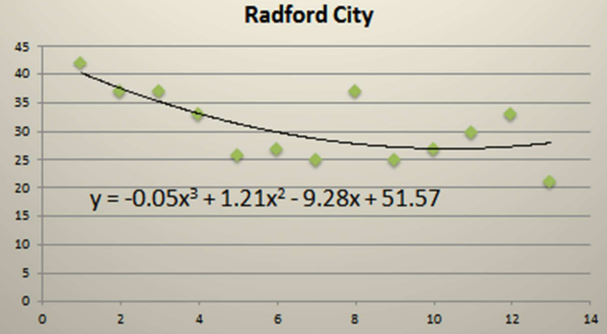
*2001 will be year 1, 2002→ 2, ...

Individual Scatterplots & Equation of Best Fit



For District 4, a 5th degree polynomial was chosen to model the data as it also had the highest correlation coefficient. This data set was harder to find an equation of best fit of modeling though as there were two obvious outliers for year 2008 and 2013 that pulled the data higher and lower, respectively.

Individual Scatterplots & Equation of Best Fit



For Radford City, there was a similar problem with year 2008 and 2013 that District 4 presented. When looking at the other data values in the set though, cubic polynomial was the best equation to fit the data.

Initial Predictions for 2020

- The State of Virginia
  - Teen Pregnancy will be at 3,200
- District 4
  - Teen Pregnancy will be at 98
- Radford City
  - Teen Pregnancy will be at 24

Actual Calculated Predictions

- State of Virginia

  $y = -76.74x^2 + 572.72x + 12,972.61$

  $y = -76.74(2016)^2 + 572.72(2016) + 12,972.61$

  $y = 2491$

- The total estimated number of teenage pregnancies for year 2016 is 2,491

Actual Calculated Predictions

- District 4

  $y = 0.01x^5 - 0.41x^4 + 4.21x^3 - 14.57x^2 - 0.92x + 284.60$

  $y = 0.01(2016)^5 - 0.41(2016)^4 + 4.21(2016)^3 - 14.57(2016)^2 - 0.92(2016) + 284.60$

  $y = -2600$

- The total estimated number of teenage pregnancies for year 2016 is -2600

Actual Calculated Predictions

- Radford City

  $y = -0.05x^3 + 1.21x^2 - 9.28x + 51.57$

  $y = -0.05(2016)^3 + 1.21(2016)^2 - 9.28(2016) + 51.57$

  $y = 8$

- The total estimated number of teenage pregnancies for year 2016 is 8

Conclusion

Based on the data sets collected for the State of Virginia, District 4, and Radford City, the rates of teenage pregnancy are decreasing. My initial predictions were 3,200 for the State of VA, 98 for District 4, and 24 for Radford City. Based on the equations of best fit though, the State of Virginia will be 2,491, District 4 will be -2600, and Radford City will be at 8. These models are not perfect as District 4 would be at zero teenage pregnancies and Radford City, which is part of District 4, will be at 8 total. There is a margin of error that needs to be accounted for and the models should not be taken as an absolute answer.