Introductory Econometrics

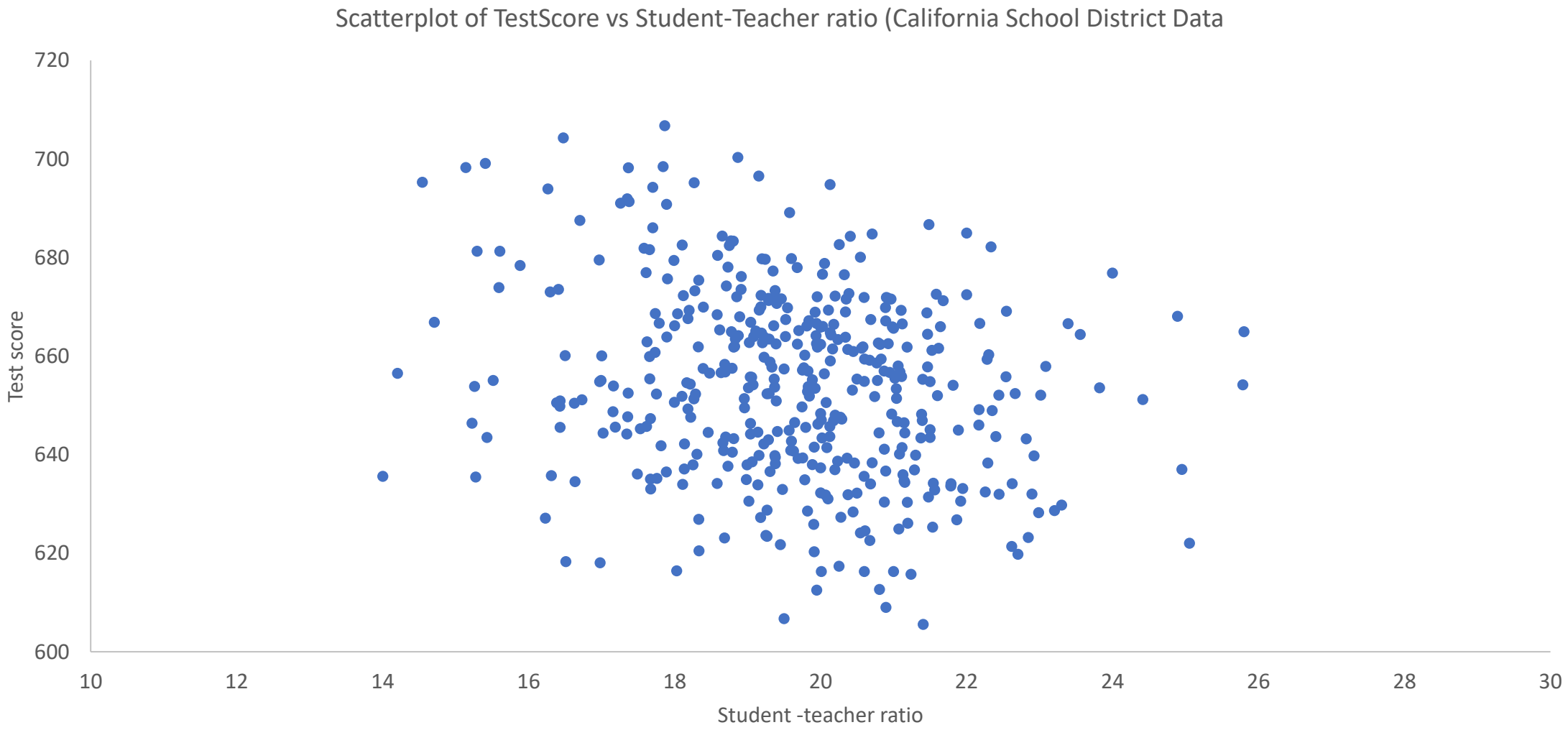
Using Excel

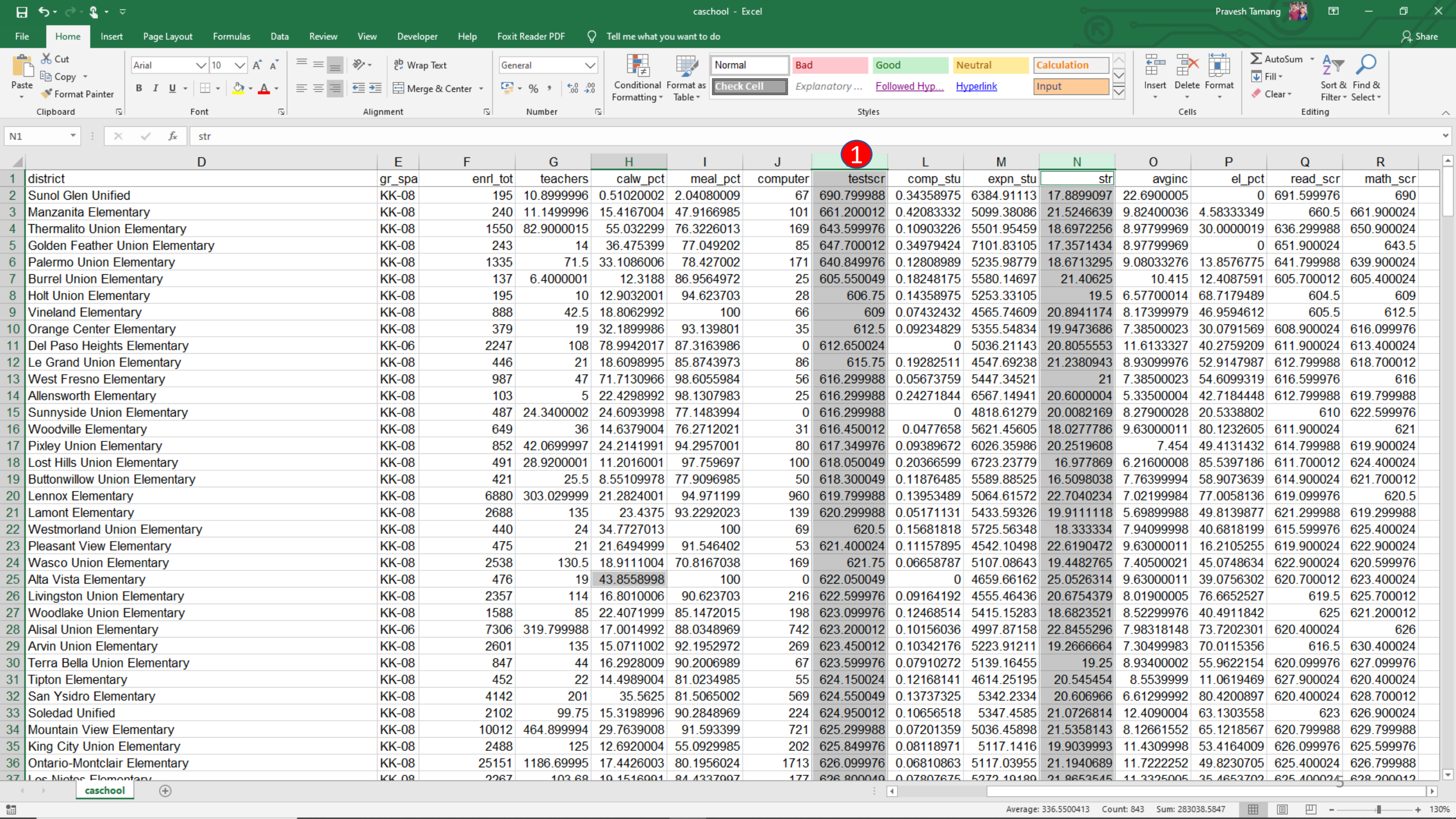
Outline

- Generate Scatter plot Figure 4.3 page 161 of Stock and Watson book.
- Generate Table 4.1 page 161 of the book.
- Linear regression: Regression Statistics, ANOVA, coefficients, p-value, standard error, confidence intervals, etc.
- Fit a trend line Figure 4.3 in page 164
- Change plot style and colour.
- Prediction
- Sensitivity of OLS to large outliers
- Exercise

Objective: Generate Figure 4.2 page 161 of the book.

You will find the data and description file here. Use the data **caschool.xlsx**.





caschool - Excel

Pravesh Tamang

File Home Insert Page Layout Formulas Data Review View Developer Help Foxit Reader PDF Tell me what you want to do Share

Paste Cut Copy Format Painter

Clipboard

Arial 10

Font

Wrap Text Merge & Center

Alignment

General

Number

Conditional Formatting Format as Table

Normal Bad Good Neutral Calculation Check Cell Explanatory ... Followed Hyp... Hyperlink Input

Styles

Insert Delete Format

Cells

AutoSum Fill Clear Sort & Filter Find & Select

Editing

N1 str

1

	D	E	F	G	H	I	J	testscr	L	M	N	O	P	Q	R
1	district	gr_spa	enrl_tot	teachers	calw_pct	meal_pct	computer	testscr	comp_stu	expn_stu	str	avginc	el_pct	read_scr	math_scr
2	Sunol Glen Unified	KK-08	195	10.8999996	0.51020002	2.04080009	67	690.799988	0.34358975	6384.91113	17.8899097	22.6900005	0	691.599976	690
3	Manzanita Elementary	KK-08	240	11.1499996	15.4167004	47.9166985	101	661.200012	0.42083332	5099.38086	21.5246639	9.82400036	4.58333349	660.5	661.900024
4	Thermalito Union Elementary	KK-08	1550	82.9000015	55.032299	76.3226013	169	643.599976	0.10903226	5501.95459	18.6972256	8.97799969	30.0000019	636.299988	650.900024
5	Golden Feather Union Elementary	KK-08	243	14	36.475399	77.049202	85	647.700012	0.34979424	7101.83105	17.3571434	8.97799969	0	651.900024	643.5
6	Palermo Union Elementary	KK-08	1335	71.5	33.1086006	78.427002	171	640.849976	0.12808989	5235.98779	18.6713295	9.08033276	13.8576775	641.799988	639.900024
7	Burrel Union Elementary	KK-08	137	6.4000001	12.3188	86.9564972	25	605.550049	0.18248175	5580.14697	21.40625	10.415	12.4087591	605.700012	605.400024
8	Holt Union Elementary	KK-08	195	10	12.9032001	94.623703	28	606.75	0.14358975	5253.33105	19.5	6.57700014	68.7179489	604.5	609
9	Vineland Elementary	KK-08	888	42.5	18.8062992	100	66	609	0.07432432	4565.74609	20.8941174	8.17399979	46.9594612	605.5	612.5
10	Orange Center Elementary	KK-08	379	19	32.1899986	93.139801	35	612.5	0.09234829	5355.54834	19.9473686	7.38500023	30.0791569	608.900024	616.099976
11	Del Paso Heights Elementary	KK-06	2247	108	78.9942017	87.3163986	0	612.650024	0	5036.21143	20.8055553	11.6133327	40.2759209	611.900024	613.400024
12	Le Grand Union Elementary	KK-08	446	21	18.6098995	85.8743973	86	615.75	0.19282511	4547.69238	21.2380943	8.93099976	52.9147987	612.799988	618.700012
13	West Fresno Elementary	KK-08	987	47	71.7130966	98.6055984	56	616.299988	0.05673759	5447.34521	21	7.38500023	54.6099319	616.599976	616
14	Allensworth Elementary	KK-08	103	5	22.4298992	98.1307983	25	616.299988	0.24271844	6567.14941	20.6000004	5.33500004	42.7184448	612.799988	619.799988
15	Sunnyside Union Elementary	KK-08	487	24.3400002	24.6093998	77.1483994	0	616.299988	0	4818.61279	20.0082169	8.27900028	20.5338802	610	622.599976
16	Woodville Elementary	KK-08	649	36	14.6379004	76.2712021	31	616.450012	0.0477658	5621.45605	18.0277786	9.63000011	80.1232605	611.900024	621
17	Pixley Union Elementary	KK-08	852	42.0699997	24.2141991	94.2957001	80	617.349976	0.09389672	6026.35986	20.2519608	7.454	49.4131432	614.799988	619.900024
18	Lost Hills Union Elementary	KK-08	491	28.9200001	11.2016001	97.759697	100	618.050049	0.20366599	6723.23779	16.977869	6.21600008	85.5397186	611.700012	624.400024
19	Buttonwillow Union Elementary	KK-08	421	25.5	8.55109978	77.9096985	50	618.300049	0.11876485	5589.88525	16.5098038	7.76399994	58.9073639	614.900024	621.700012
20	Lennox Elementary	KK-08	6880	303.029999	21.2824001	94.971199	960	619.799988	0.13953489	5064.61572	22.7040234	7.02199984	77.0058136	619.099976	620.5
21	Lamont Elementary	KK-08	2688	135	23.4375	93.2292023	139	620.299988	0.05171131	5433.59326	19.9111118	5.69899988	49.8139877	621.299988	619.299988
22	Westmorland Union Elementary	KK-08	440	24	34.7727013	100	69	620.5	0.15681818	5725.56348	18.333334	7.94099998	40.6818199	615.599976	625.400024
23	Pleasant View Elementary	KK-08	475	21	21.6494999	91.546402	53	621.400024	0.11157895	4542.10498	22.6190472	9.63000011	16.2105255	619.900024	622.900024
24	Wasco Union Elementary	KK-08	2538	130.5	18.9111004	70.8167038	169	621.75	0.06658787	5107.08643	19.4482765	7.40500021	45.0748634	622.900024	620.599976
25	Alta Vista Elementary	KK-08	476	19	43.8558998	100	0	622.050049	0	4659.66162	25.0526314	9.63000011	39.0756302	620.700012	623.400024
26	Livingston Union Elementary	KK-08	2357	114	16.8010006	90.623703	216	622.599976	0.09164192	4555.46436	20.6754379	8.01900005	76.6652527	619.5	625.700012
27	Woodlake Union Elementary	KK-08	1588	85	22.4071999	85.1472015	198	623.099976	0.12468514	5415.15283	18.6823521	8.52299976	40.4911842	625	621.200012
28	Alisal Union Elementary	KK-06	7306	319.799988	17.0014992	88.0348969	742	623.200012	0.10156036	4997.87158	22.8455296	7.98318148	73.7202301	620.400024	626
29	Arvin Union Elementary	KK-08	2601	135	15.0711002	92.1952972	269	623.450012	0.10342176	5223.91211	19.2666664	7.30499983	70.0115356	616.5	630.400024
30	Terra Bella Union Elementary	KK-08	847	44	16.2928009	90.2006989	67	623.599976	0.07910272	5139.16455	19.25	8.93400002	55.9622154	620.099976	627.099976
31	Tipton Elementary	KK-08	452	22	14.4989004	81.0234985	55	624.150024	0.12168141	4614.25195	20.545454	8.5539999	11.0619469	627.900024	620.400024
32	San Ysidro Elementary	KK-08	4142	201	35.5625	81.5065002	569	624.550049	0.13737325	5342.2334	20.606966	6.61299992	80.4200897	620.400024	628.700012
33	Soledad Unified	KK-08	2102	99.75	15.3198996	90.2848969	224	624.950012	0.10656518	5347.4585	21.0726814	12.4090004	63.1303558	623	626.900024
34	Mountain View Elementary	KK-08	10012	464.899994	29.7639008	91.593399	721	625.299988	0.07201359	5036.45898	21.5358143	8.12661552	65.1218567	620.799988	629.799988
35	King City Union Elementary	KK-08	2488	125	12.6920004	55.0929985	202	625.849976	0.08118971	5117.1416	19.9039993	11.4309998	53.4164009	626.099976	625.599976
36	Ontario-Montclair Elementary	KK-08	25151	1186.69995	17.4426003	80.1956024	1713	626.099976	0.06810863	5117.03955	21.1940689	11.7222252	49.8230705	625.400024	626.799988
37	Los Nietos Elementary	KK-08	2267	103.68	19.1516991	84.4337997	177	626.800049	0.07807675	5272.19189	21.8653545	11.3325005	35.4653702	625.400024	628.200012

caschool

Average: 336.5500413 Count: 843 Sum: 283038.5847 130%

caschool - Excel

Pravesh Tamang

File Home Insert Page Layout Formulas Data Review View Developer Help Foxit Reader PDF Tell me what you want to do Share

PivotTable Recommended PivotTables Table — Tables
Pictures Online Pictures Shapes SmartArt Screenshot — Illustrations
Get Add-ins My Add-ins Bing Maps People Graph — Add-ins
Recommended Charts PivotChart 3D Map — Charts
Line Column Win/ Loss — Sparklines
Slicer Timeline — Filters
Link — Links
Text Box Header & Footer WordArt Signature Line Object — Text
Equation Symbol — Symbols

B1 str

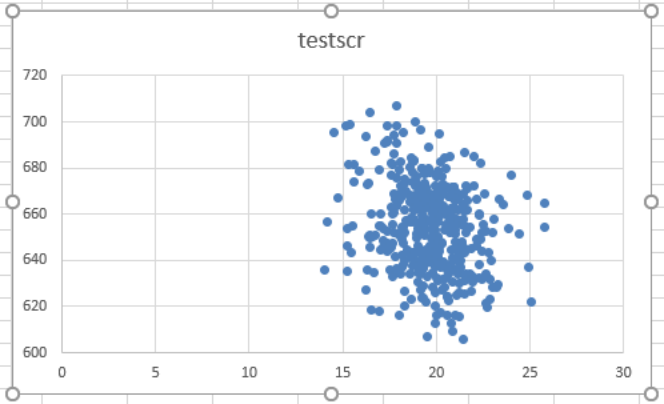
Scatter

Bubble

More Scatter Charts...

3

	A	B	C
1		str	testscr
2		17.88991	690.8
3		21.52466	661.2
4		18.69723	643.6
5		17.35714	647.7
6		18.67133	640.85
7		21.40625	605.55
8		19.5	606.75
9		20.89412	609
10		19.94737	612.5
11		20.80556	612.65
12		21.23809	615.75
13		21	616.3
14		20.6	616.3
15		20.00822	616.3
16		18.02778	616.45
17		20.25196	617.35
18		16.97787	618.05
19		16.5098	618.3
20		22.70402	619.8
21		19.91111	620.3
22		18.33333	620.5
23		22.61905	621.4
24		19.44828	621.75
25		25.05263	622.05
26		20.67544	622.6
27		18.68235	623.1
28		22.84553	623.2
29		19.26667	623.45
30		19.25	623.6
31		20.54545	624.15
32		20.60697	624.55
33		21.07268	624.95
34		21.53581	625.3
35		19.904	625.85
36		21.19407	626.1
37		21.86535	626.8
38		18.32965	626.9
39		16.22857	627.1
40		19.17857	627.25
41		20.27737	627.3
42		22.98614	628.25
43		20.44444	628.4
44		19.82085	628.55
45		23.20522	628.65
46		19.26697	628.75
47		23.30189	629.8
48		21.18829	630.35



Format Shape

Fill

Line

Make sure ***str*** is in the 1st column

2

caschool Sheet3

Average: 336.8984867 Count: 842 Sum: 282994.7288 100%

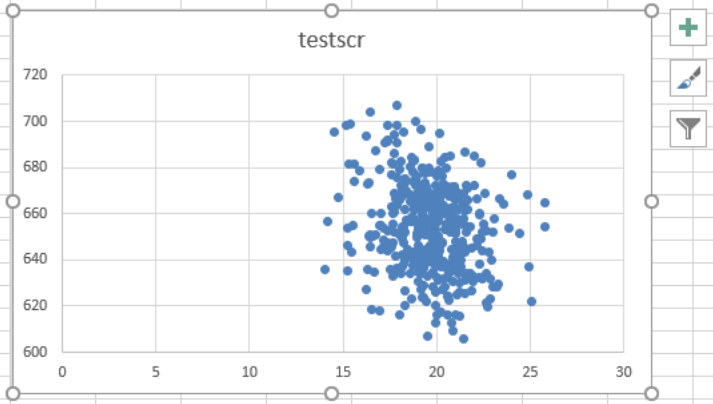
Chart Tools | caschool - Excel | Pravesh Tamang

File Home Insert Page Layout Formulas Data Review View Developer Help Foxit Reader PDF Design Format Tell me what you want to do Share

Add Chart Element | Quick Layout | Change Colors | Chart Layouts | Chart Styles | Switch Row/ Column | Select Data | Data | Change Chart Type | Type | Move Chart | Location

Chart 3

	A	B	C
1		str	testscr
2		17.88991	690.8
3		21.52466	661.2
4		18.69723	643.6
5		17.35714	647.7
6		18.67133	640.85
7		21.40625	605.55
8		19.5	606.75
9		20.89412	609
10		19.94737	612.5
11		20.80556	612.65
12		21.23809	615.75
13		21	616.3
14		20.6	616.3
15		20.00822	616.3
16		18.02778	616.45
17		20.25196	617.35
18		16.97787	618.05
19		16.5098	618.3
20		22.70402	619.8
21		19.91111	620.3
22		18.33333	620.5
23		22.61905	621.4
24		19.44828	621.75
25		25.05263	622.05
26		20.67544	622.6
27		18.68235	623.1
28		22.84553	623.2
29		19.26667	623.45
30		19.25	623.6
31		20.54545	624.15
32		20.60697	624.55
33		21.07268	624.95
34		21.53581	625.3
35		19.904	625.85
36		21.19407	626.1
37		21.86535	626.8
38		18.32965	626.9
39		16.22857	627.1
40		19.17857	627.25
41		20.27737	627.3
42		22.98614	628.25
43		20.44444	628.4
44		19.82085	628.55
45		23.20522	628.65
46		19.26697	628.75
47		23.30189	629.8



4

Move Chart

Choose where you want the chart to be placed:

New sheet: Figure 4.2

5

Object in: Sheet3

OK Cancel

Format Chart Area

Chart Options | Text Options

Fill

Border

caschool | Sheet3

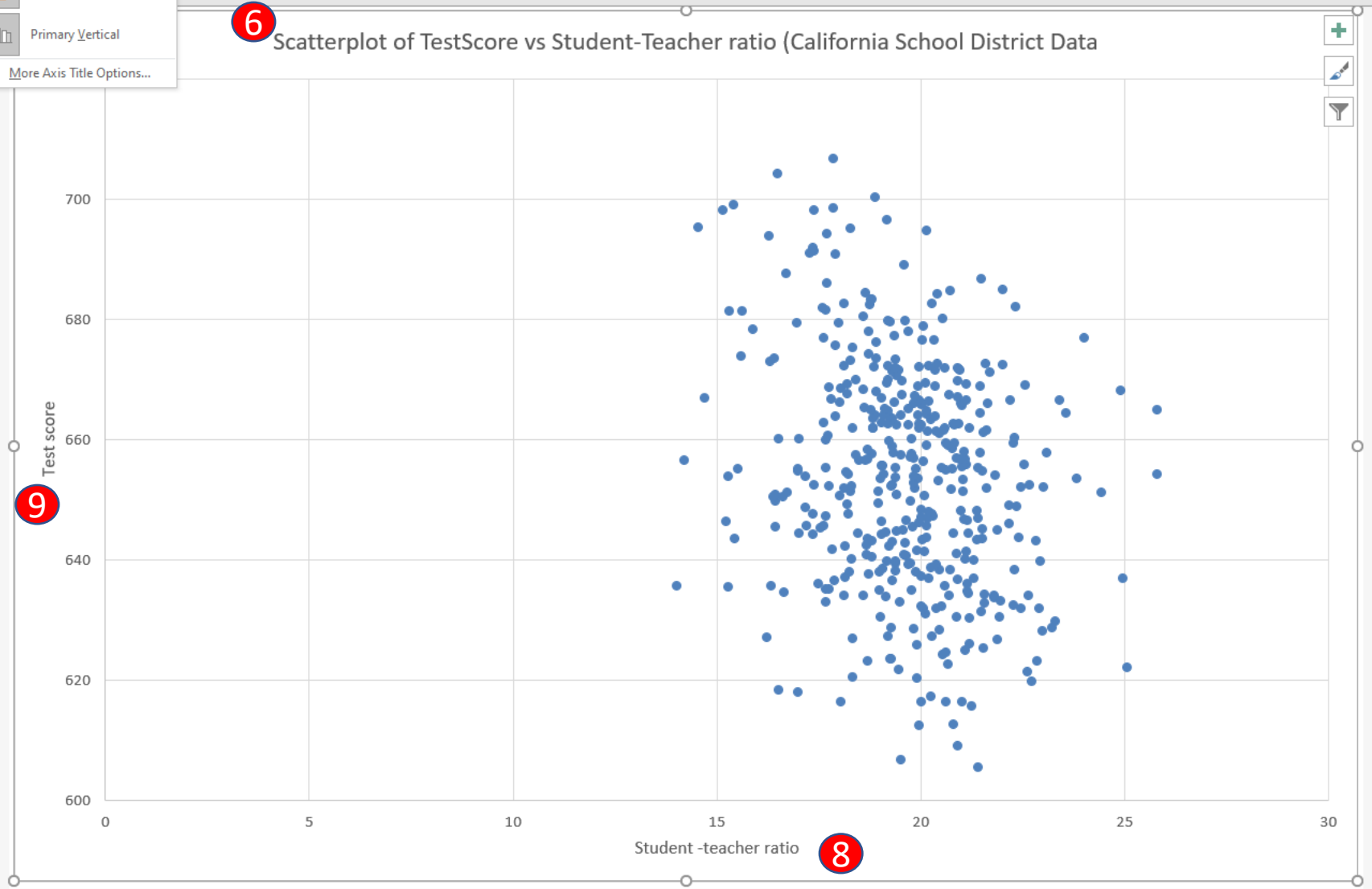
Average: 336.8984867 Count: 842 Sum: 282994.7288 100%

Add Chart Element | Quick Layout | Change Colors | Chart Styles | Switch Row/ Column | Select Data | Data | Change Chart Type | Type | Move Chart | Location

7

- Axes
- Axis Titles
  - Primary Horizontal
  - Primary Vertical
  - More Axis Title Options...
- Chart Title
- Data Labels
- Data Table
- Error Bars
- Gridlines
- Legend
- Lines
- Trendline
- Up/Down Bars

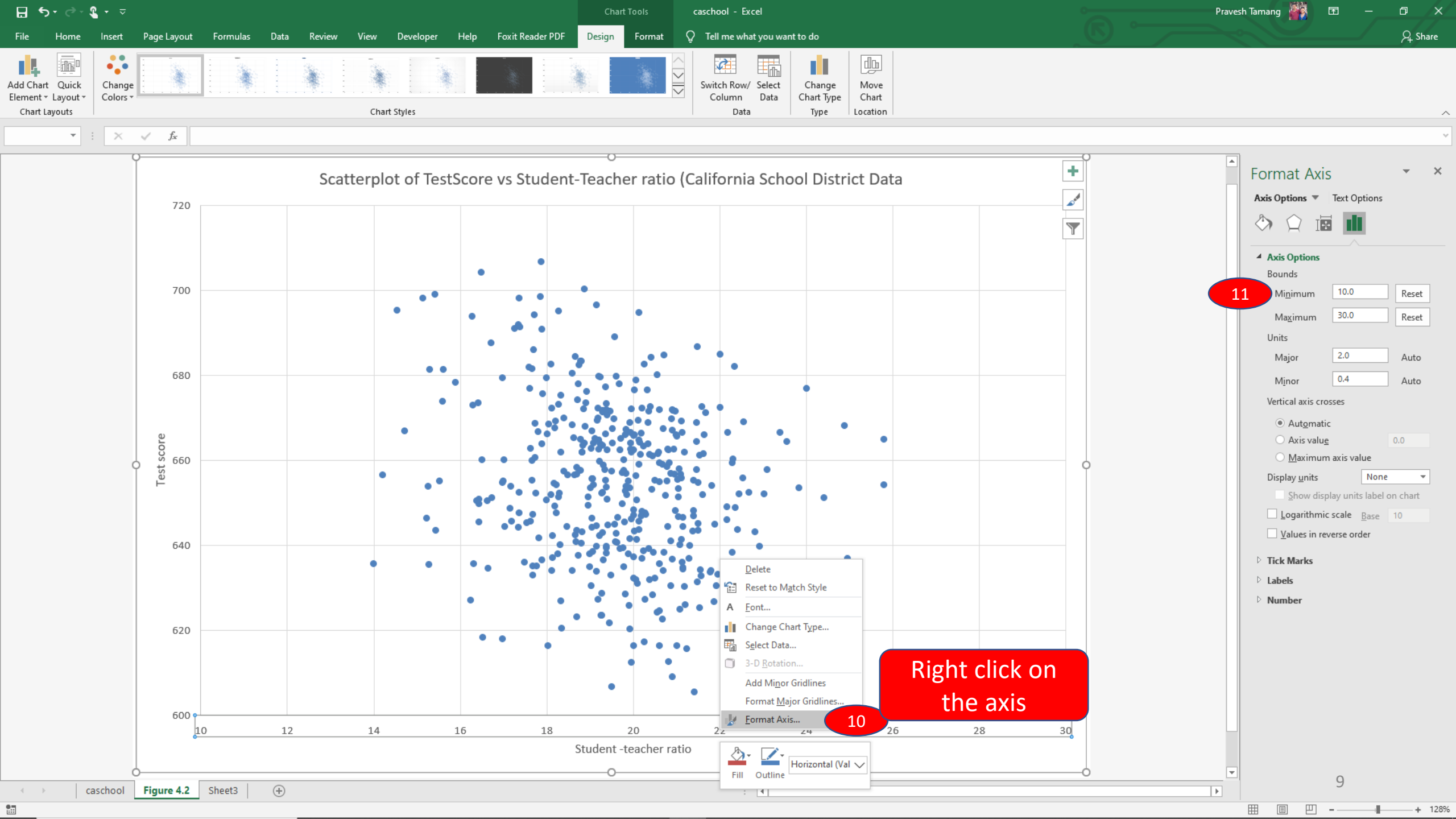
6



9

8

Format Chart Area
Chart Options | Text Options
Fill
Border


caschool - Excel
Chart Tools
Pravesh Tamang
File Home Insert Page Layout Formulas Data Review View Developer Help Foxit Reader PDF Design Format
Tell me what you want to do
Share
Add Chart Element
Quick Layout
Change Colors
Chart Layouts
Chart Styles
Switch Row/ Column
Select Data
Data
Change Chart Type
Type
Move Chart
Location
Scatterplot of TestScore vs Student-Teacher ratio (California School District Data
Test score
720
700
680
660
640
620
600
10
12
14
16
18
20
22
24
26
28
30
Student -teacher ratio
Delete
Reset to Match Style
Font...
Change Chart Type...
Select Data...
3-D Rotation...
Add Minor Gridlines
Format Major Gridlines...
Format Axis...
10
Right click on the axis
Fill
Outline
Horizontal (Val
Format Axis
Axis Options
Text Options
Axis Options
Bounds
11
Minimum
10.0
Reset
Maximum
30.0
Reset
Units
Major
2.0
Auto
Minor
0.4
Auto
Vertical axis crosses
Automatic
Axis value
0.0
Maximum axis value
Display units
None
Show display units label on chart
Logarithmic scale
Base
10
Values in reverse order
Tick Marks
Labels
Number
caschool
Figure 4.2
Sheet3
9
128%

Objective: Generate Table 4.1 page 161 of the book.

Objective: Generate Table 4.1 page 161 of the book.

			Percentile					
	Average	Standard Deviation	10%	25%	50% (Median)	60%	75%	90%
Student Teacher Ratio	19.64	1.88	17.34	18.58	19.72	20.07	20.87	21.86
Test Score	654.15	19.03	630.39	640.05	654.44	659.4	666.66	678.86

Cut Copy Format Painter Paste — Clipboard

Arial 10 — Font

Wrap Text Merge & Center — Alignment

General — Number

Conditional Formatting Format as Table

Normal Bad Good Neutral Calculation
Check Cell Explanatory ... Followed Hyp... Hyperlink Input — Styles

Insert Delete Format — Cells

AutoSum Fill Clear Sort & Filter Find & Select — Editing

N12

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1		str	testscr																								
2		17.88991	690.8																								
3		21.52466	661.2																								
4		18.69723	643.6						Percentile																		
5		17.35714	647.7				Average	Standard Deviation	10%	25%	50% (Median)	60%	75%	90%													
6		18.67133	640.85			Student Teacher Ratio	19.64043	1.889558288	17.3486	18.58236	19.72320843	20.0783	20.87181	21.8674													
7		21.40625	605.55			Test Score	654.1565	19.03065157	630.395	640.05	654.4499817	659.4	666.6625	678.86													
8		19.5	606.75																								
9		20.89412	609																								
10		19.94737	612.5																								
11		20.80556	612.65																								
12		21.23809	615.75																								
13		21	616.3																								
14		20.6	616.3																								
15		20.00822	616.3																								
16		18.02778	616.45																								
17		20.25196	617.35																								
18		16.97787	618.05																								
19		16.5098	618.3																								
20		22.70402	619.8																								
21		19.91111	620.3																								
22		18.33333	620.5																								
23		22.61905	621.4																								
24		19.44828	621.75																								
25		25.05263	622.05																								
26		20.67544	622.6																								
27		18.68235	623.1																								
28		22.84553	623.2																								
29		19.26667	623.45																								
30		19.25	623.6																								
31		20.54545	624.15																								
32		20.60697	624.55																								
33		21.07268	624.95																								
34		21.53581	625.3																								
35		19.904	625.85																								
36		21.19407	626.1																								
37		21.86535	626.8																								
38		18.32965	626.9																								
39		16.22857	627.1																								
40		19.17857	627.25																								
41		20.27737	627.3																								
42		22.98614	628.25																								
43		20.44444	628.4																								
44		19.82085	628.55																								
45		23.20522	628.65																								
46		19.26697	628.75																								
47		23.30189	629.8																								
48		21.18829	630.35																								

Formulae used:

Cell G6=AVERAGE(B2:B421)
Cell H6=STDEV.P(B2:B421)
Cell I6=PERCENTILE.INC(B2:B421, 0.1)
Cell J6=PERCENTILE.INC(B2:B421, 0.25)
Cell K6=PERCENTILE.INC(B2:B421, 0.5)
Cell L6=PERCENTILE.INC(B2:B421, 0.6)
Cell M6=PERCENTILE.INC(B2:B421, 0.75)
Cell N6=PERCENTILE.INC(B2:B421, 0.9)

Repeat the same for G7:GN7

caschool Figure 4.2 Sheet3

100%

Linear regression

Regression Statistics, ANOVA, coefficients, p-value, standard error, confidence intervals, etc.

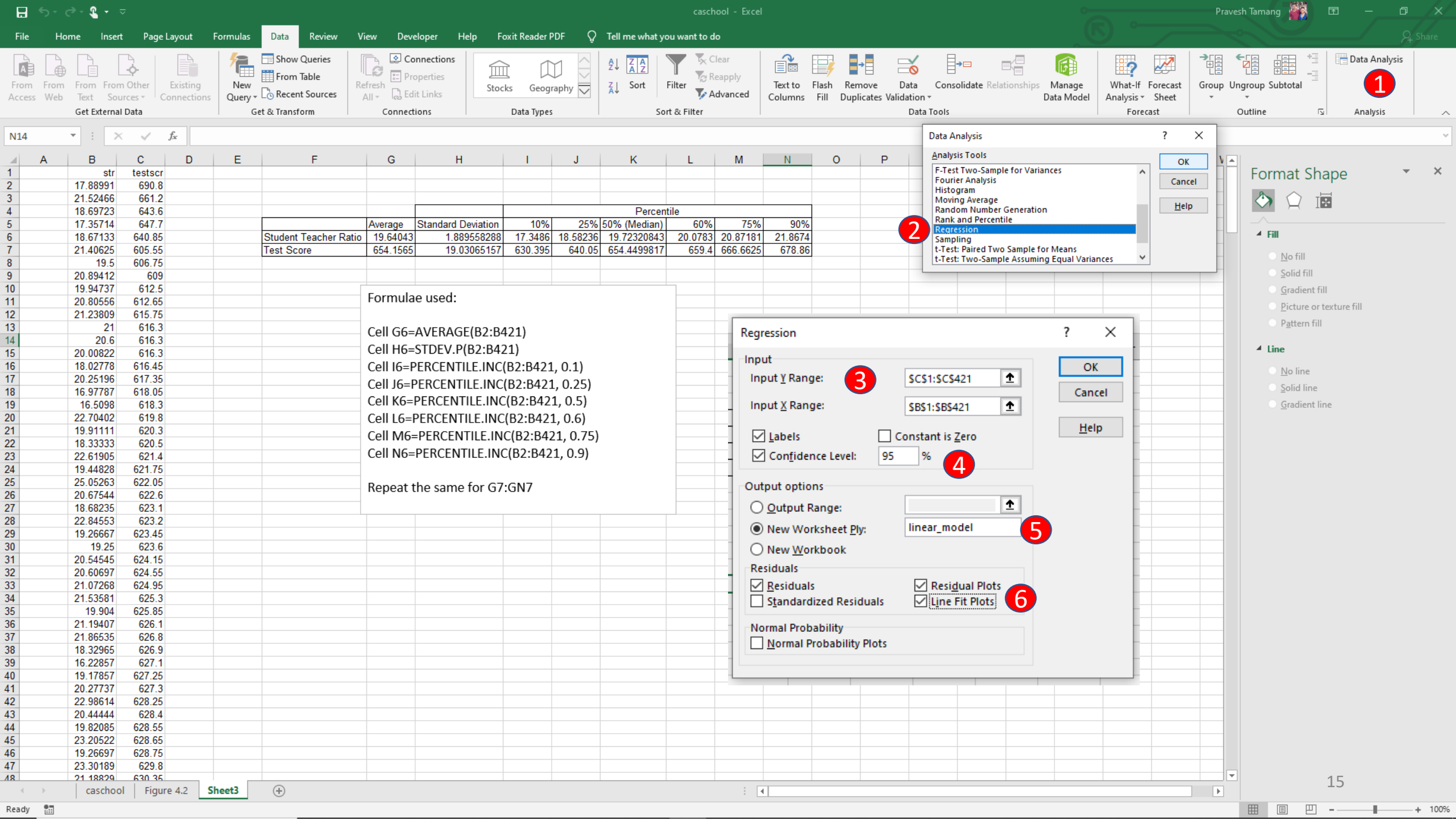
Estimate the regression coefficents in

$$\hat{TestScore} = \hat{\beta}_0 + \hat{\beta}_1 STR$$

Where, $TestScore$ is the average test score in the district, and STR is the student-teacher ratio.

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$\hat{\beta}_o = \bar{Y} - \hat{\beta}_1 \bar{X}$$



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		str	testscr													
2		17.88991	690.8													
3		21.52466	661.2													
4		18.69723	643.6						Percentile							
5		17.35714	647.7				Average	Standard Deviation	10%	25%	50% (Median)	60%	75%	90%		
6		18.67133	640.85			Student Teacher Ratio	19.64043	1.889558288	17.3486	18.58236	19.72320843	20.0783	20.87181	21.8674		
7		21.40625	605.55			Test Score	654.1565	19.03065157	630.395	640.05	654.4499817	659.4	666.6625	678.86		
8		19.5	606.75													
9		20.89412	609													
10		19.94737	612.5													
11		20.80556	612.65													
12		21.23809	615.75													
13		21	616.3													
14		20.6	616.3													
15		20.00822	616.3													
16		18.02778	616.45													
17		20.25196	617.35													
18		16.97787	618.05													
19		16.5098	618.3													
20		22.70402	619.8													
21		19.91111	620.3													
22		18.33333	620.5													
23		22.61905	621.4													
24		19.44828	621.75													
25		25.05263	622.05													
26		20.67544	622.6													
27		18.68235	623.1													
28		22.84553	623.2													
29		19.26667	623.45													
30		19.25	623.6													
31		20.54545	624.15													
32		20.60697	624.55													
33		21.07268	624.95													
34		21.53581	625.3													
35		19.904	625.85													
36		21.19407	626.1													
37		21.86535	626.8													
38		18.32965	626.9													
39		16.22857	627.1													
40		19.17857	627.25													
41		20.27737	627.3													
42		22.98614	628.25													
43		20.44444	628.4													
44		19.82085	628.55													
45		23.20522	628.65													
46		19.26697	628.75													
47		23.30189	629.8													

Formulae used:

Cell G6=AVERAGE(B2:B421)
Cell H6=STDEV.P(B2:B421)
Cell I6=PERCENTILE.INC(B2:B421, 0.1)
Cell J6=PERCENTILE.INC(B2:B421, 0.25)
Cell K6=PERCENTILE.INC(B2:B421, 0.5)
Cell L6=PERCENTILE.INC(B2:B421, 0.6)
Cell M6=PERCENTILE.INC(B2:B421, 0.75)
Cell N6=PERCENTILE.INC(B2:B421, 0.9)

Repeat the same for G7:GN7

caschool - Excel

Pravesh Tamang

File Home Insert Page Layout Formulas Data Review View Developer Help Foxit Reader PDF Tell me what you want to do Share

Cut Copy Format Painter Paste Clipboard

Arial 10 Font

Wrap Text Merge & Center Alignment

General Number

Conditional Formatting Format as Table

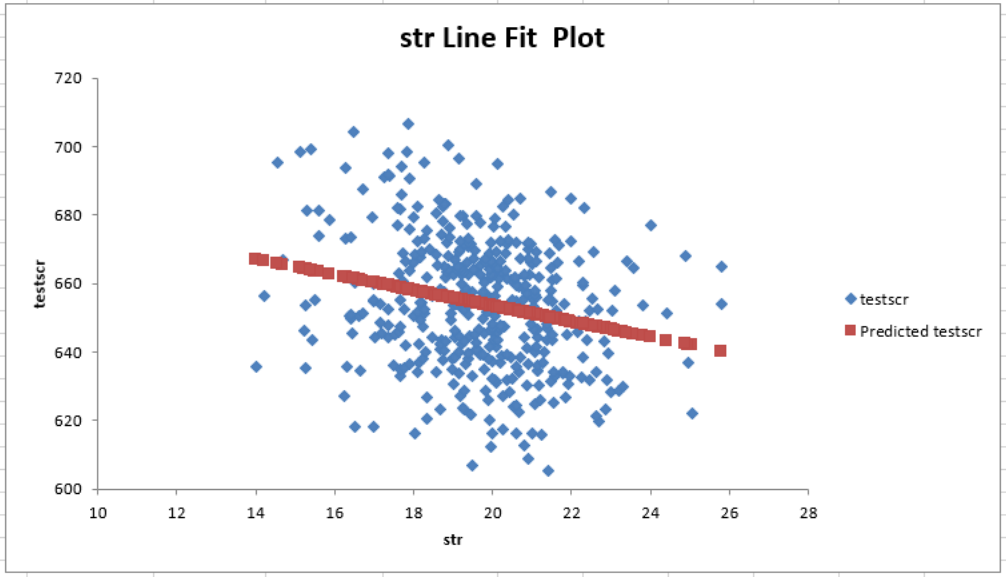
Normal Bad Good Neutral Calculation Check Cell Explanatory ... Followed Hyp... Hyperlink Input Styles

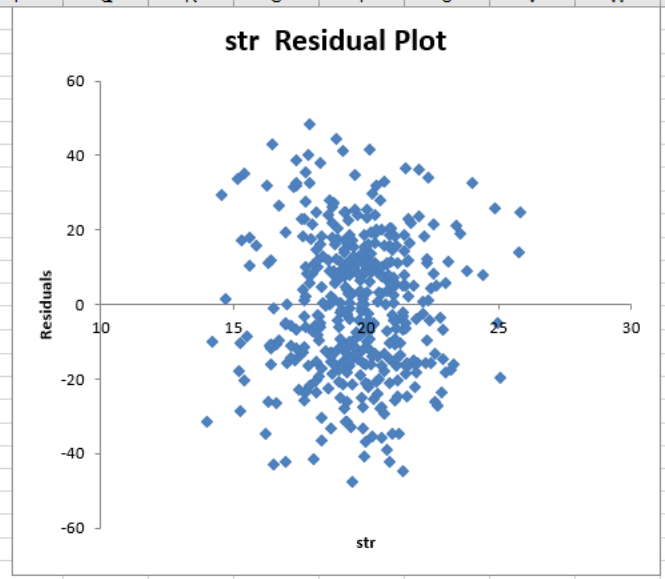
Insert Delete Format Cells

AutoSum Fill Clear Sort & Filter Find & Select Editing

I10

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	*Regression Statistics*								
4	Multiple R	0.226362751							
5	R Square	0.051240095							
6	Adjusted R Square	0.048970335							
7	Standard Error	18.5809675							
8	Observations	420							
9									
10	ANOVA								
11		*df*	*SS*	*MS*	*F*	*Significance F*			
12	Regression	1	7794.110041	7794.110041	22.57511055	0.0000028			
13	Residual	418	144315.4836	345.2523531					
14	Total	419	152109.5936						
15									
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
17	Intercept	698.9329523	9.467491444	73.82451375	0.0000000	680.323126	717.5427786	680.323126	717.5427786
18	str	-2.279808287	0.479825567	-4.751327241	0.0000028	-3.222980034	-1.33663654	-3.222980034	-1.33663654
19									
20									
21									
22	RESIDUAL OUTPUT								
23									
24	*Observation*	*Predicted testscr*	*Residuals*						
25	1	658.1473878	32.65259995						
26	2	649.8608451	11.33916707						
27	3	656.3068625	-12.70688694						
28	4	659.361993	-11.66198075						
29	5	656.3659006	-15.51592502						
30	6	650.1308062	-44.58075736						
31	7	654.4766907	-47.72669073						
32	8	651.2983704	-42.29837043						
33	9	653.456776	-40.95677604						
34	10	651.5002748	-38.85025043						
35	11	650.5141689	-34.76416887						
36	12	651.0569783	-34.75699051						
37	13	651.9689007	-35.66891295						
38	14	653.3180537	-37.01806593						
39	15	657.8330732	-41.38306101						
40	16	652.7623644	-35.41238879						
41	17	660.2266658	-42.17661698						
42	18	661.2937649	-42.99371604						
43	19	647.1721317	-27.37214393						
44	20	653.5394346	-33.23944678						
45	21	657.1364656	-36.63646562						
46	22	647.3658612	-25.96583674						
47	23	654.5946103	-32.84461035						





caschool Figure 4.2 linear_model Sheet3

Ready 100%

Cut Copy Format Painter Paste Clipboard
Arial 10 Font
Wrap Text Merge & Center Alignment
General Number
Conditional Formatting Format as Table
Normal Bad Good Neutral Calculation Check Cell Explanatory ... Followed Hyp... Hyperlink Input Styles
Insert Delete Format Cells
AutoSum Fill Clear Sort & Filter Find & Select Editing

E22

	A	B	C	D	E	F	G	H	I	J	K
2											
3	*Regression Statistics*										
4	Multiple R	0.226362751									
5	R Square	0.051240095									
6	Adjusted R Square	0.048970335									
7	Standard Error	18.5809675									
8	Observations	420									
9											
10	ANOVA										
11		*df*	*SS*	*MS*	*F*	*Significance F*					
12	Regression	1	7794.110041	7794.110041	22.57511055	0.0000028					
13	Residual	418	144315.4836	345.2523531							
14	Total	419	152109.5936								
15											
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*		
17	Intercept	698.9329523	9.467491444	73.82451375	0.0000000	680.323126	717.542779	680.323126	717.5427786		
18	str	-2.279808287	0.479825567	-4.75132724	0.0000028	-3.222980034	-1.33663654	-3.22298003	-1.33663654		
19											
20											
21											
22	RESIDUAL OUTPUT										
23											
24	*Observation*	*Predicted testscr*	*Residuals*								
25	1	658.1473878	32.65259995								

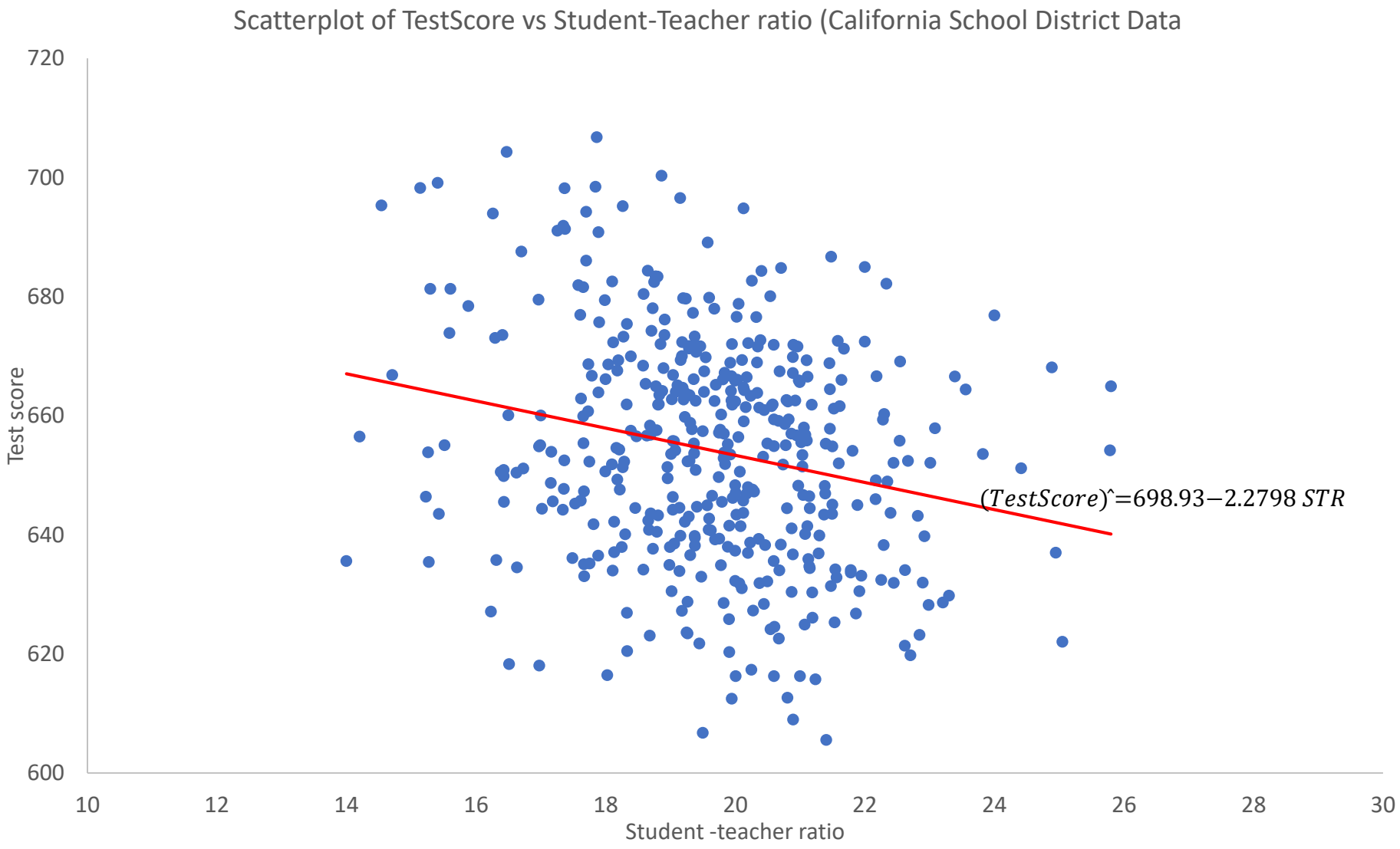
Plot Area

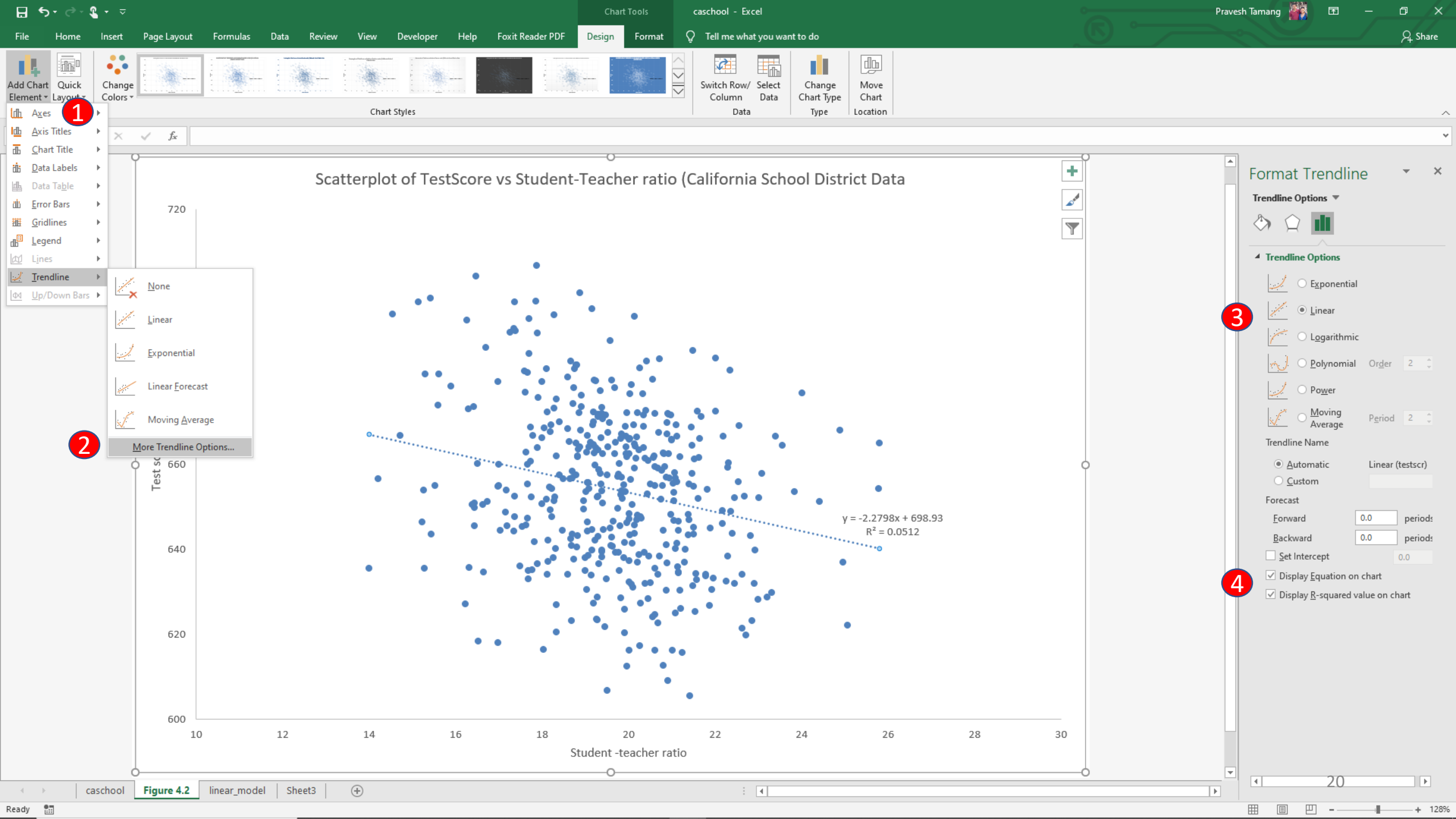
caschool Figure 4.2 linear_model Sheet3

Ready 190%

Fit a trend line Figure 4.3 in page 164

Fit a trend line Figure 4.3 in page 164




caschool - Excel
Chart Tools
Pravesh Tamang
File Home Insert Page Layout Formulas Data Review View Developer Help Foxit Reader PDF Design Format
Tell me what you want to do
Share
Add Chart Element
Quick Layout
Change Colors
Chart Styles
Switch Row/ Column
Select Data
Data
Change Chart Type
Type
Move Chart
Location
Axes
Axis Titles
Chart Title
Data Labels
Data Table
Error Bars
Gridlines
Legend
Lines
Trendline
Up/Down Bars
None
Linear
Exponential
Linear Forecast
Moving Average
More Trendline Options...
1
2
Scatterplot of TestScore vs Student-Teacher ratio (California School District Data
720
660
640
620
600
Test sc
10
12
14
16
18
20
22
24
26
28
30
Student -teacher ratio
y = -2.2798x + 698.93
R² = 0.0512
Format Trendline
Trendline Options
Trendline Options
Exponential
Linear
Logarithmic
Polynomial
Order
Power
Moving Average
Period
3
Trendline Name
Automatic
Linear (testscr)
Custom
Forecast
Forward
0.0
periods
Backward
0.0
periods
Set Intercept
0.0
Display Equation on chart
Display R-squared value on chart
4
caschool
Figure 4.2
linear_model
Sheet3
Ready
20
128%

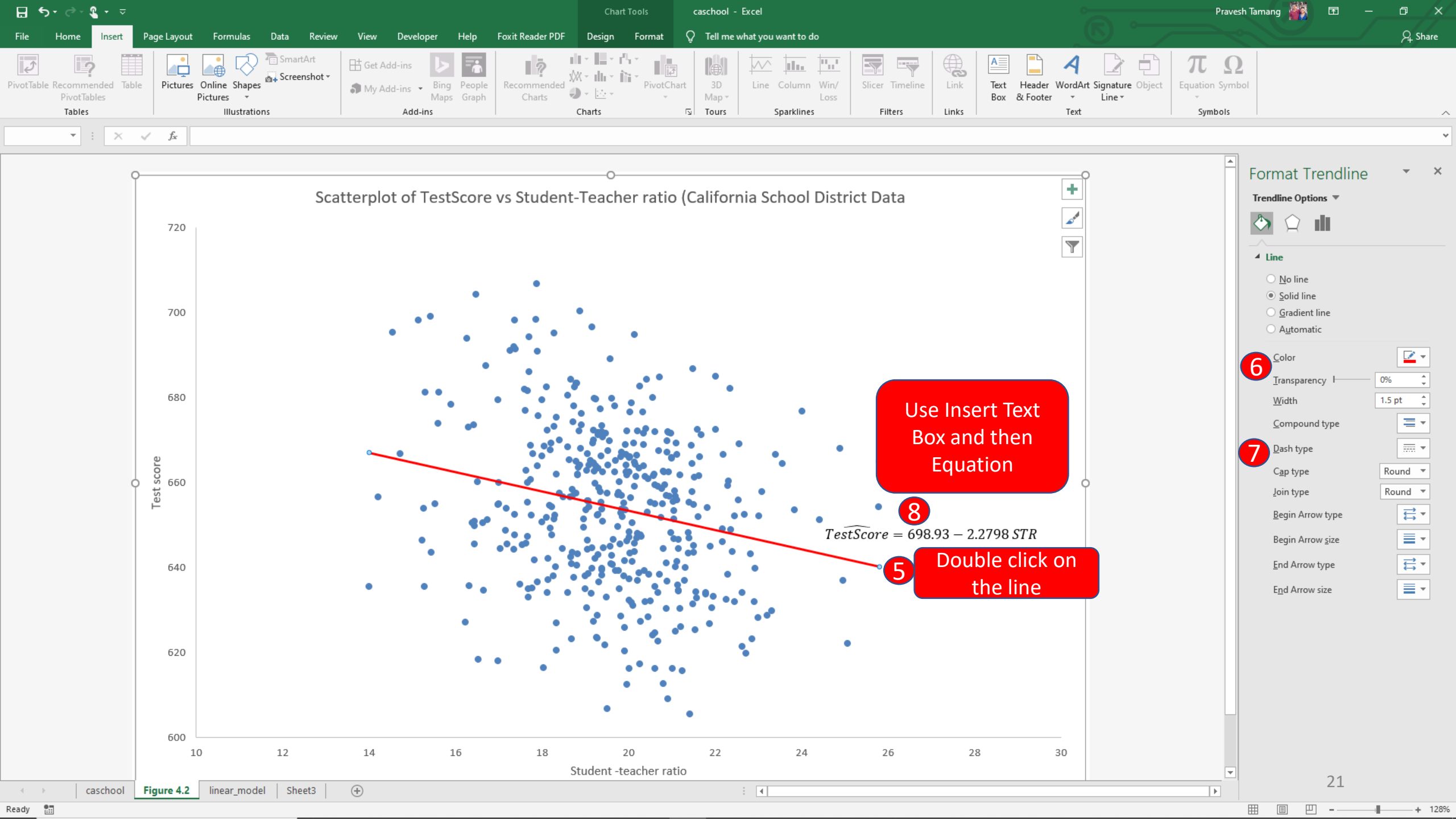
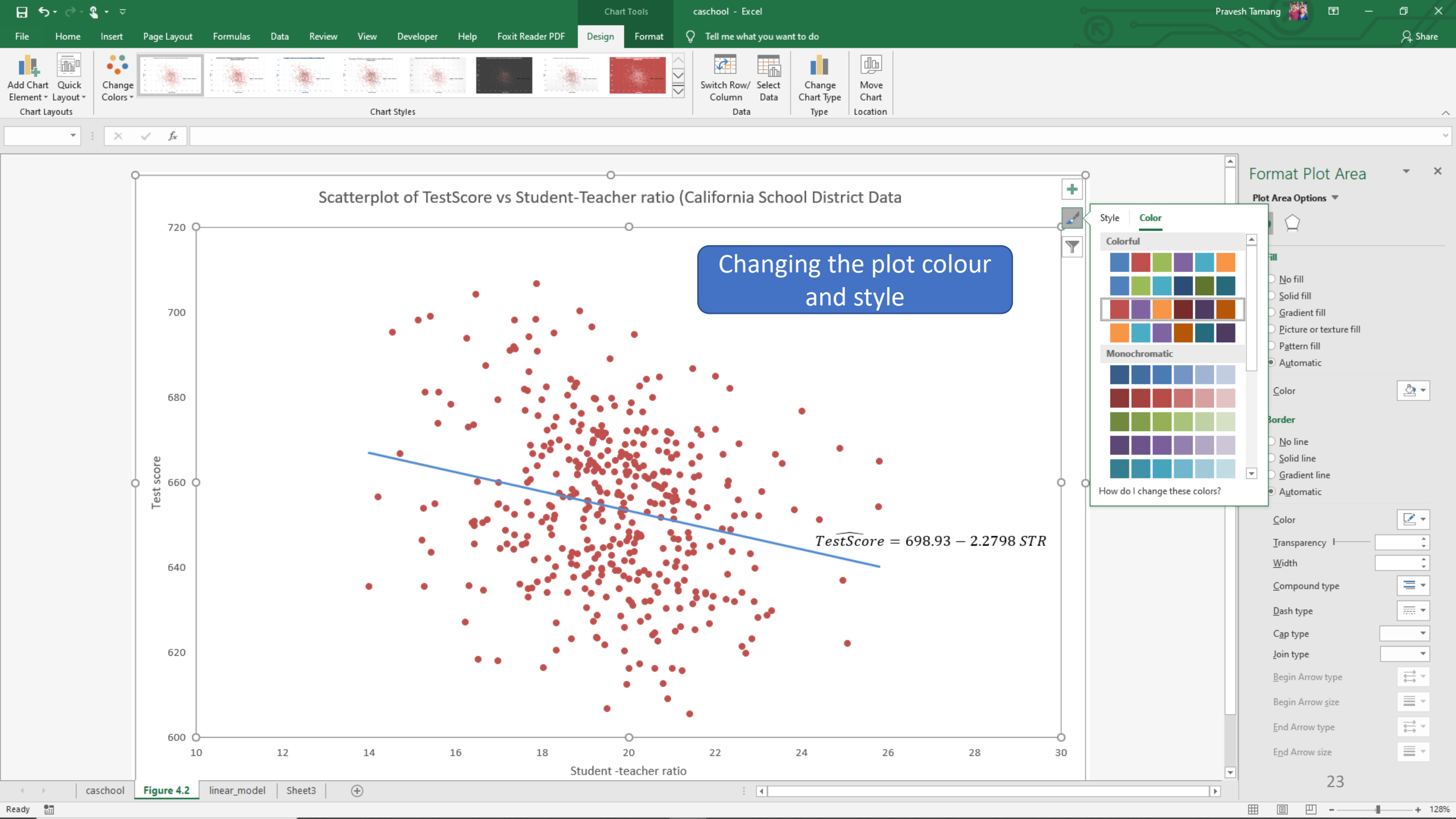


Chart Tools
caschool - Excel
Pravesh Tamang
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Developer
Help
Foxit Reader PDF
Design
Format
Tell me what you want to do
Share
Scatterplot of TestScore vs Student-Teacher ratio (California School District Data
Test score
Student -teacher ratio
720
700
680
660
640
620
600
10
12
14
16
18
20
22
24
26
28
30
Use Insert Text Box and then Equation
8
$\widehat{TestScore} = 698.93 - 2.2798\ STR$
5
Double click on the line
Format Trendline
Trendline Options
Line
No line
Solid line
Gradient line
Automatic
6
Color
Transparency
0%
Width
1.5 pt
Compound type
7
Dash type
Cap type
Round
Join type
Round
Begin Arrow type
Begin Arrow size
End Arrow type
End Arrow size
caschool
Figure 4.2
linear_model
Sheet3
Ready
21
128%

Changing the plot colour and style


caschool - Excel
Chart Tools
Pravesh Tamang
File Home Insert Page Layout Formulas Data Review View Developer Help Foxit Reader PDF Design Format
Tell me what you want to do
Share
Add Chart Element
Quick Layout
Change Colors
Chart Layouts
Chart Styles
Switch Row/ Column
Select Data
Data
Change Chart Type
Type
Move Chart
Location
Scatterplot of TestScore vs Student-Teacher ratio (California School District Data
Changing the plot colour and style
$\widehat{TestScore} = 698.93 - 2.2798\ STR$
Test score
Student -teacher ratio
720
700
680
660
640
620
600
10
12
14
16
18
20
22
24
26
28
30
Style
Color
Colorful
Monochromatic
How do I change these colors?
Format Plot Area
Plot Area Options
Fill
No fill
Solid fill
Gradient fill
Picture or texture fill
Pattern fill
Automatic
Color
Border
No line
Solid line
Gradient line
Automatic
Color
Transparency
Width
Compound type
Dash type
Cap type
Join type
Begin Arrow type
Begin Arrow size
End Arrow type
End Arrow size
caschool
Figure 4.2
linear_model
Sheet3
Ready
128%

Prediction

Syntax

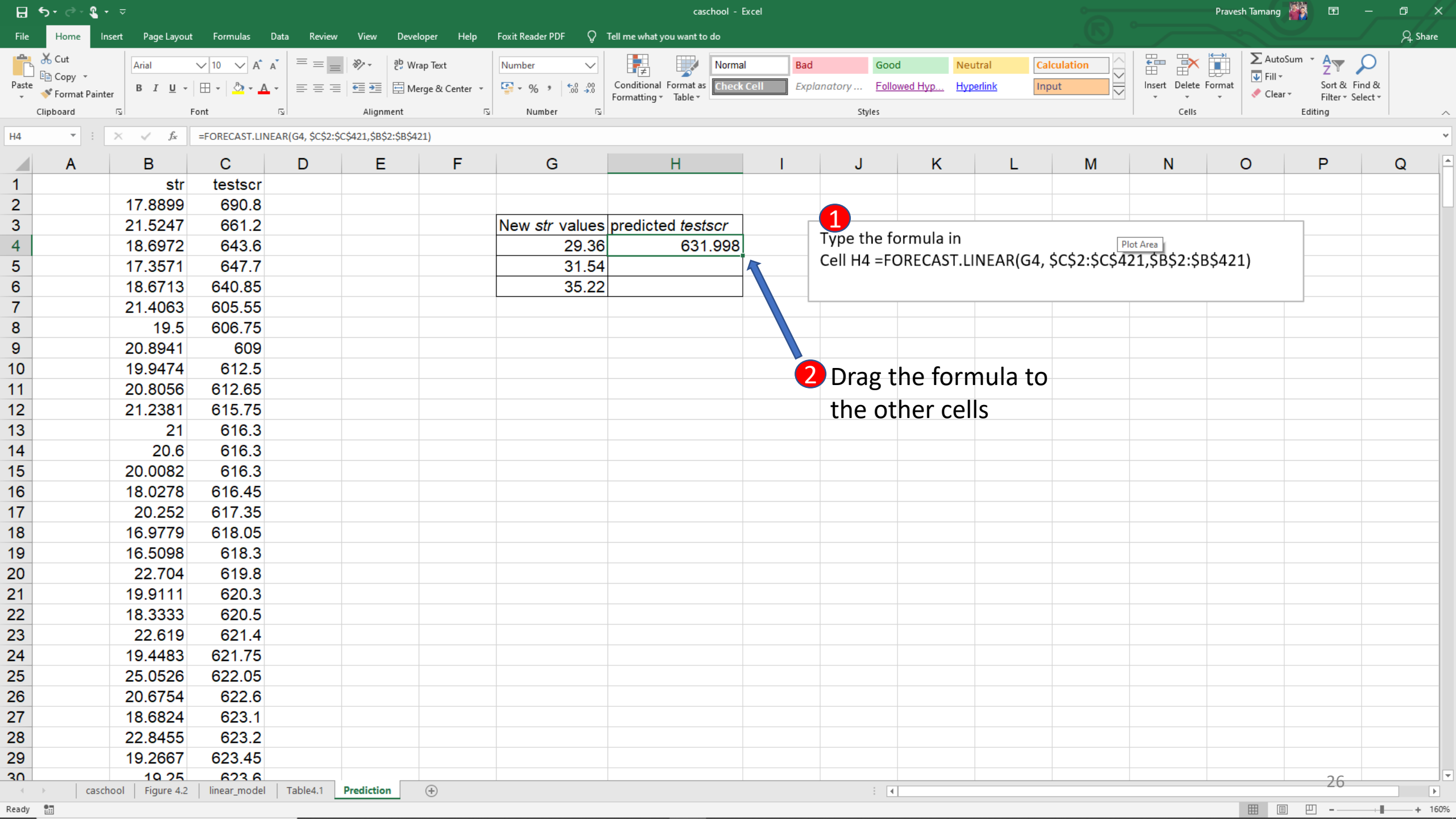
FORECAST.LINEAR(x, known_y's, known_x's)

- or -

FORECAST(x, known_y's, known_x's)

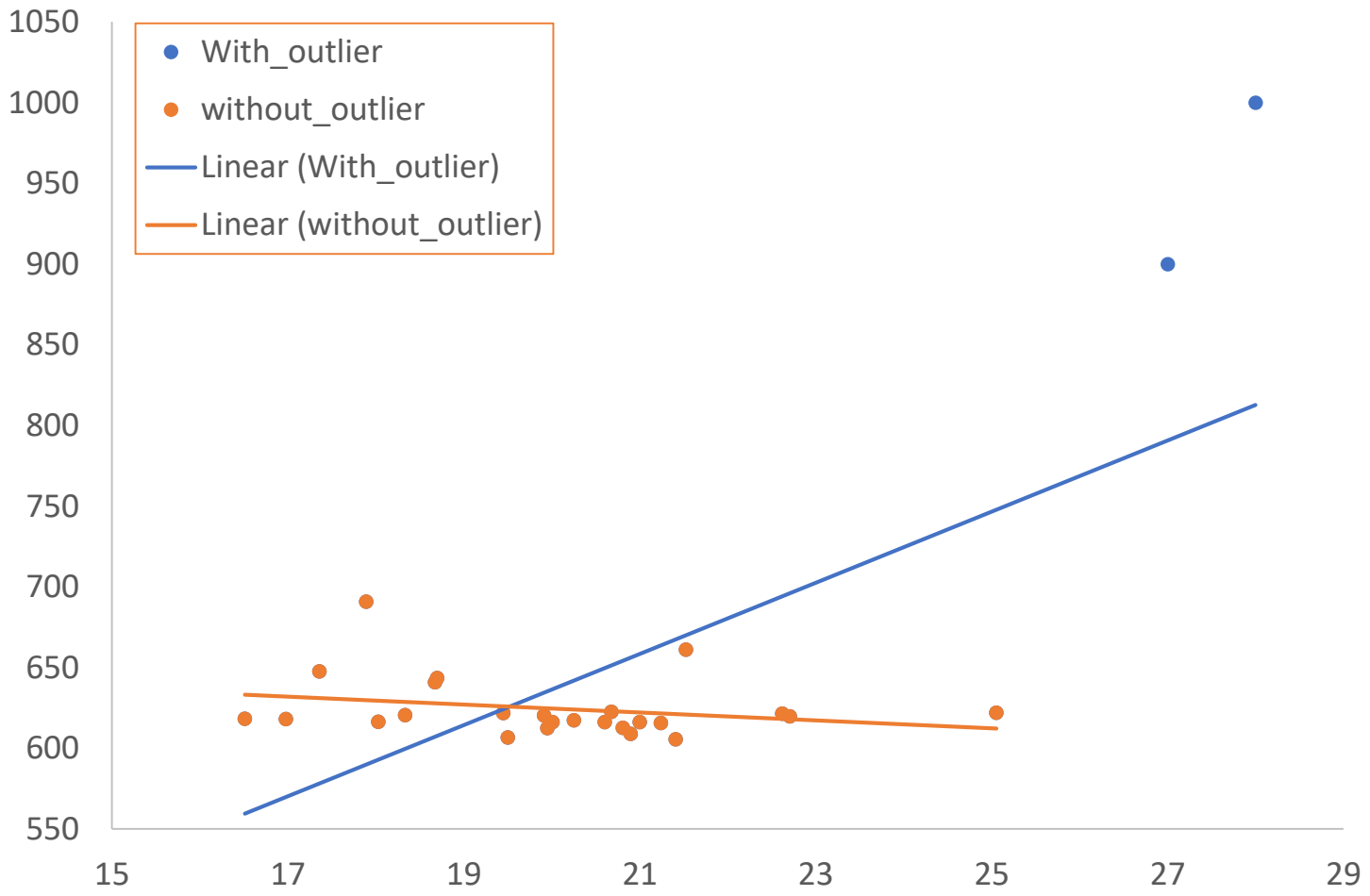
The FORECAST/FORECAST.LINEAR function syntax has the following arguments:

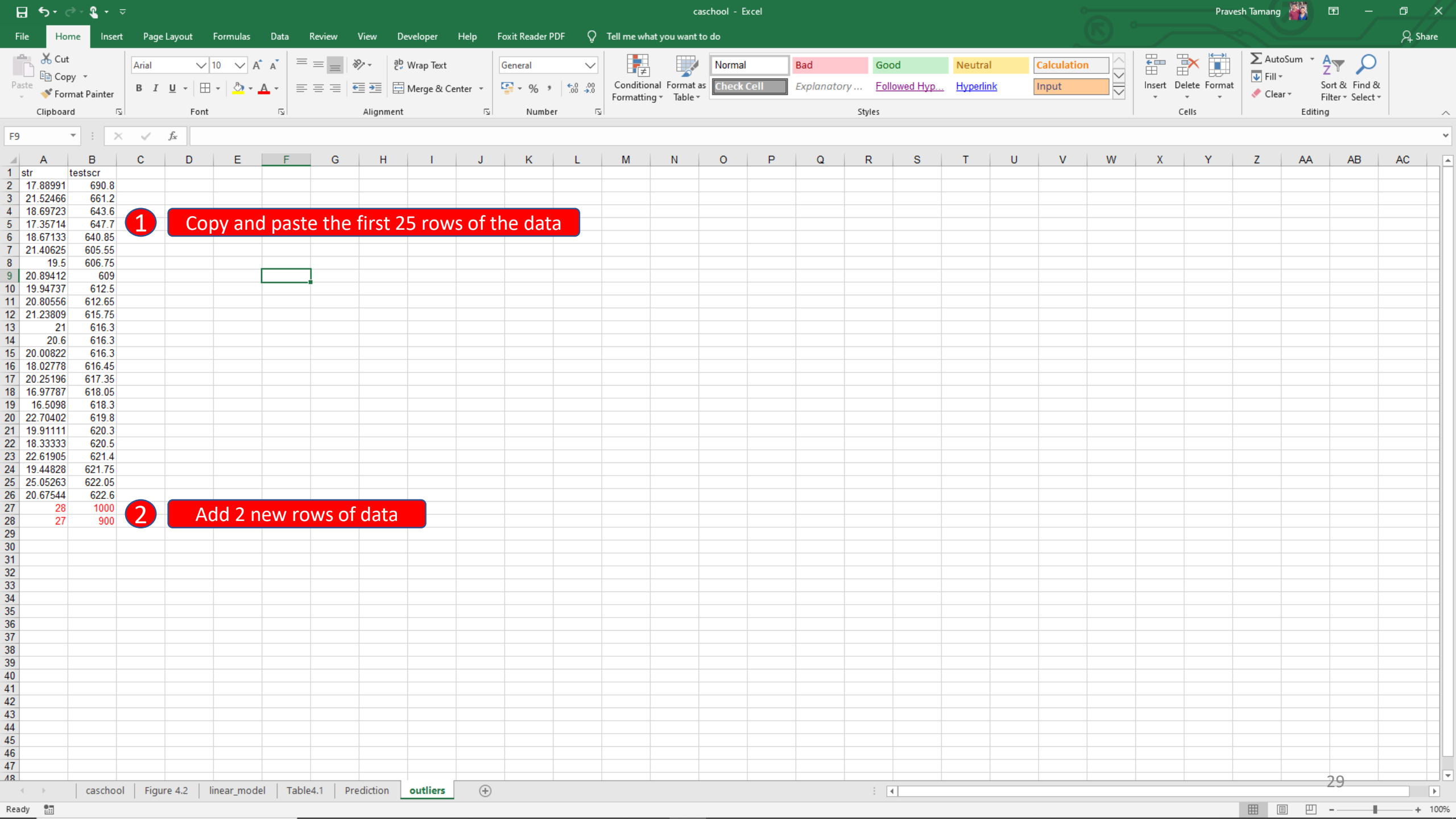
Argument	Required	Refers to
x	yes	The data point for which you want to predict a value.
known_y's	yes	The dependent array or range of data.
known_x's	yes	The independent array or range of data.



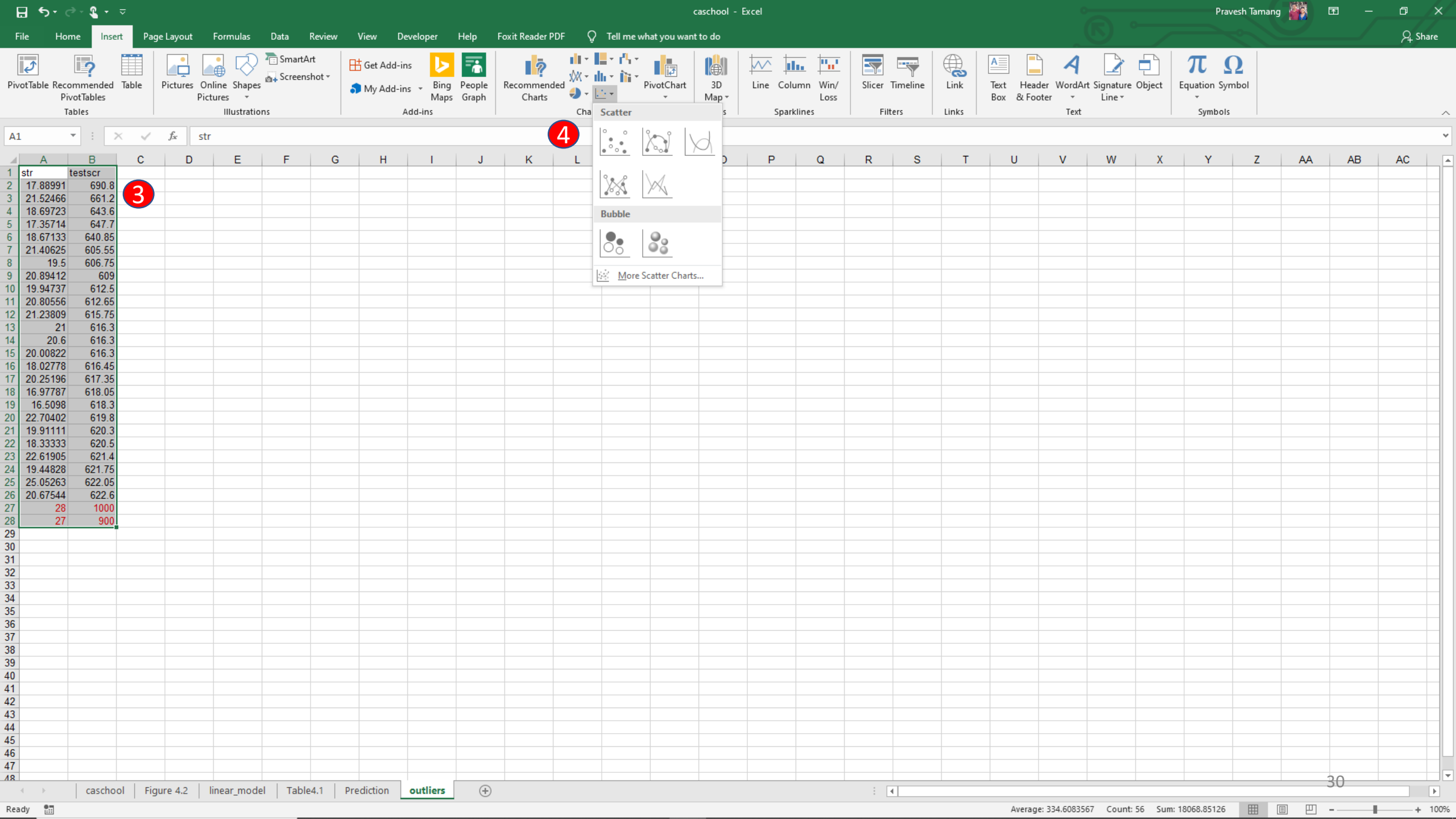
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		str	testscr														
2		17.8899	690.8														
3		21.5247	661.2				New *str* values	predicted *testscr*									
4		18.6972	643.6				29.36	631.998									
5		17.3571	647.7				31.54										
6		18.6713	640.85				35.22										
7		21.4063	605.55														
8		19.5	606.75														
9		20.8941	609														
10		19.9474	612.5														
11		20.8056	612.65														
12		21.2381	615.75														
13		21	616.3														
14		20.6	616.3														
15		20.0082	616.3														
16		18.0278	616.45														
17		20.252	617.35														
18		16.9779	618.05														
19		16.5098	618.3														
20		22.704	619.8														
21		19.9111	620.3														
22		18.3333	620.5														
23		22.619	621.4														
24		19.4483	621.75														
25		25.0526	622.05														
26		20.6754	622.6														
27		18.6824	623.1														
28		22.8455	623.2														
29		19.2667	623.45														
30		19.25	623.6														

The sensitivity of OLS to large outliers


With_outlier
without_outlier
Linear (With_outlier)
Linear (without_outlier)
1050
1000
950
900
850
800
750
700
650
600
550
15
17
19
21
23
25
27
29



	A	B
1	str	testscr
2	17.88991	690.8
3	21.52466	661.2
4	18.69723	643.6
5	17.35714	647.7
6	18.67133	640.85
7	21.40625	605.55
8	19.5	606.75
9	20.89412	609
10	19.94737	612.5
11	20.80556	612.65
12	21.23809	615.75
13	21	616.3
14	20.6	616.3
15	20.00822	616.3
16	18.02778	616.45
17	20.25196	617.35
18	16.97787	618.05
19	16.5098	618.3
20	22.70402	619.8
21	19.91111	620.3
22	18.33333	620.5
23	22.61905	621.4
24	19.44828	621.75
25	25.05263	622.05
26	20.67544	622.6
27	28	1000
28	27	900



str	testscr
17.88991	690.8
21.52466	661.2
18.69723	643.6
17.35714	647.7
18.67133	640.85
21.40625	605.55
19.5	606.75
20.89412	609
19.94737	612.5
20.80556	612.65
21.23809	615.75
21	616.3
20.6	616.3
20.00822	616.3
18.02778	616.45
20.25196	617.35
16.97787	618.05
16.5098	618.3
22.70402	619.8
19.91111	620.3
18.33333	620.5
22.61905	621.4
19.44828	621.75
25.05263	622.05
20.67544	622.6
28	1000
27	900

Chart Tools | caschool - Excel | Pravesh Tamang

File | Home | Insert | Page Layout | Formulas | Data | Review | View | Developer | Help | Foxit Reader PDF | Design | Format | Tell me what you want to do | Share

Add Chart Element | Quick Layout | Change Colors | Chart Layouts | Chart Styles | Switch Row/ Column | Select Data | Change Chart Type | Move Chart | Data | Location

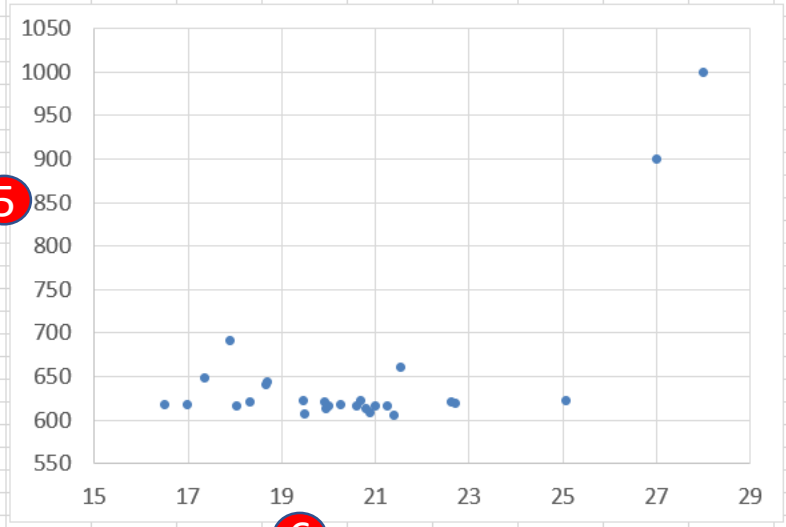
7

Select Data

Change the data range included in the chart.

A1

	A	B
1	str	testscr
2	17.88991	690.8
3	21.52466	661.2
4	18.69723	643.6
5	17.35714	647.7
6	18.67133	640.85
7	21.40625	605.55
8	19.5	606.75
9	20.89412	609
10	19.94737	612.5
11	20.80556	612.65
12	21.23809	615.75
13	21	616.3
14	20.6	616.3
15	20.00822	616.3
16	18.02778	616.45
17	20.25196	617.35
18	16.97787	618.05
19	16.5098	618.3
20	22.70402	619.8
21	19.91111	620.3
22	18.33333	620.5
23	22.61905	621.4
24	19.44828	621.75
25	25.05263	622.05
26	20.67544	622.6
27	28	1000
28	27	900



5

Adjust the limits of both the axis

6

Select Data Source

Chart data range: =outliers!$A$1:$B$28

Switch Row/Column

Legend Entries (Series)

8

Add | Edit | Remove

testscr

Horizontal (Category) Axis Labels

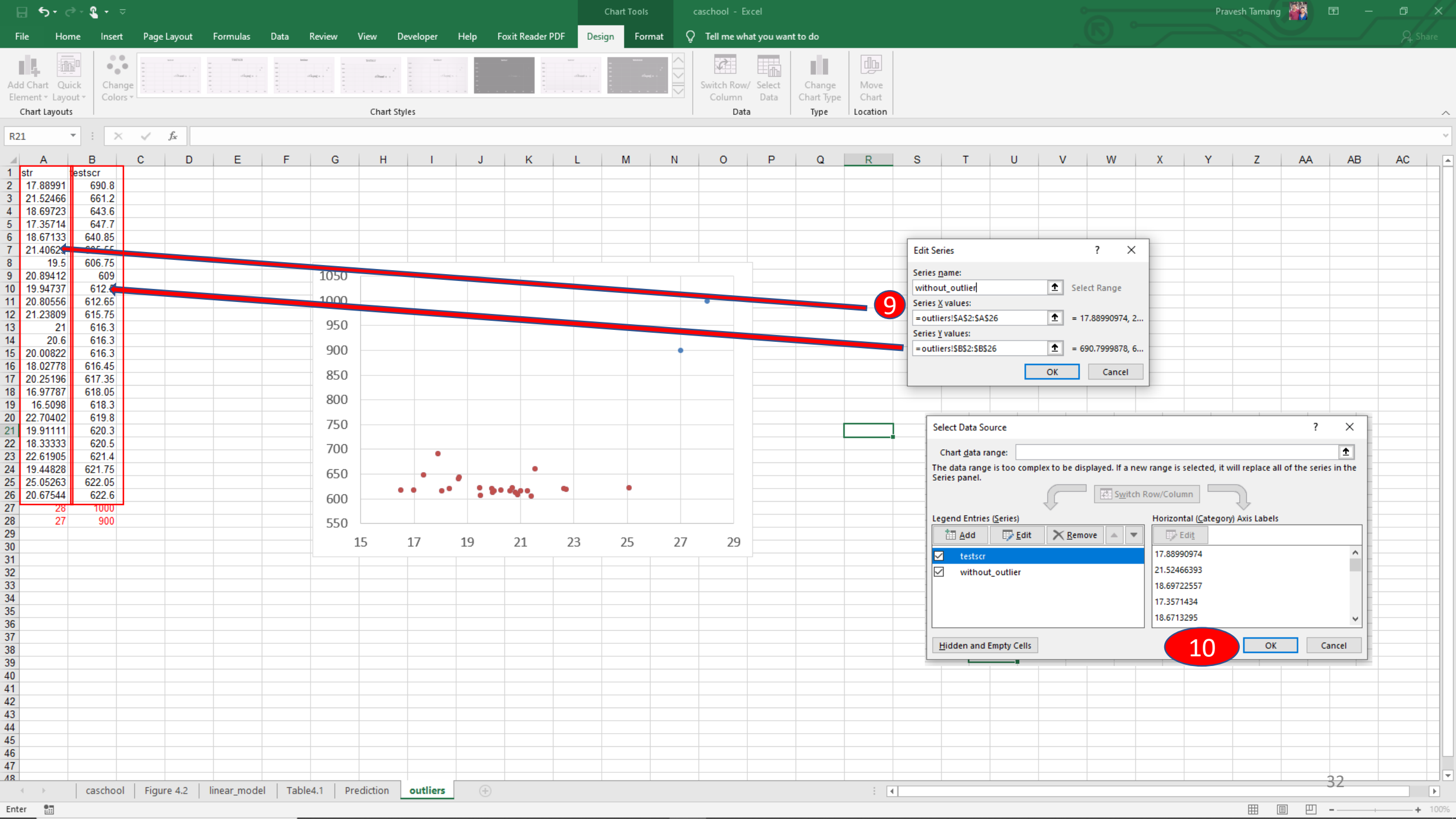
Edit

17.88990974
21.52466393
18.69722557
17.3571434
18.6713295

Hidden and Empty Cells | OK | Cancel

caschool | Figure 4.2 | linear_model | Table4.1 | Prediction | outliers

Point


caschool - Excel
Chart Tools
Pravesh Tamang
File Home Insert Page Layout Formulas Data Review View Developer Help Foxit Reader PDF Design Format
Tell me what you want to do
Add Chart Element
Quick Layout
Change Colors
Chart Layouts
Chart Styles
Switch Row/ Column
Select Data
Data
Change Chart Type
Type
Move Chart
Location
R21
str testscr
17.88991 690.8
21.52466 661.2
18.69723 643.6
17.35714 647.7
18.67133 640.85
19.5 606.75
20.89412 609
19.94737 612.
20.80556 612.65
21.23809 615.75
21 616.3
20.6 616.3
20.00822 616.3
18.02778 616.45
20.25196 617.35
16.97787 618.05
16.5098 618.3
22.70402 619.8
19.91111 620.3
18.33333 620.5
22.61905 621.4
19.44828 621.75
25.05263 622.05
20.67544 622.6
28 1000
27 900
1050
1000
950
900
850
800
750
700
650
600
550
15 17 19 21 23 25 27 29
Edit Series
Series name:
without_outlier
Select Range
Series X values:
=outliers!$A$2:$A$26
= 17.88990974, 2...
Series Y values:
=outliers!$B$2:$B$26
= 690.7999878, 6...
OK
Cancel
9
Select Data Source
Chart data range:
The data range is too complex to be displayed. If a new range is selected, it will replace all of the series in the Series panel.
Switch Row/Column
Legend Entries (Series)
Add
Edit
Remove
testscr
without_outlier
Horizontal (Category) Axis Labels
Edit
17.88990974
21.52466393
18.69722557
17.3571434
18.6713295
Hidden and Empty Cells
10
OK
Cancel
caschool
Figure 4.2
linear_model
Table4.1
Prediction
outliers
Enter
32

Add Chart Element | Quick Layout | Change Colors | Chart Styles | Switch Row/ Column | Select Data | Change Chart Type | Move Chart
Chart Layouts | Data | Type | Location

Chart 2
=SERIES(outliers!$B$1,outliers!$A$2:$A$28,outliers!$B$2:$B$28,1)

	A	B
1	str	testscr
2	17.88991	690.8
3	21.52466	661.2
4	18.69723	643.6
5	17.35714	647.7
6	18.67133	640.85
7	21.40625	605.55
8	19.5	606.75
9	20.89412	609
10	19.94737	612.5
11	20.80556	612.65
12	21.23809	615.75
13	21	616.3
14	20.6	616.3
15	20.00822	616.3
16	18.02778	616.45
17	20.25196	617.35
18	16.97787	618.05
19	16.5098	618.3
20	22.70402	619.8
21	19.91111	620.3
22	18.33333	620.5
23	22.61905	621.4
24	19.44828	621.75
25	25.05263	622.05
26	20.67544	622.6
27	28	1000
28	27	900

Click on the blue point

11

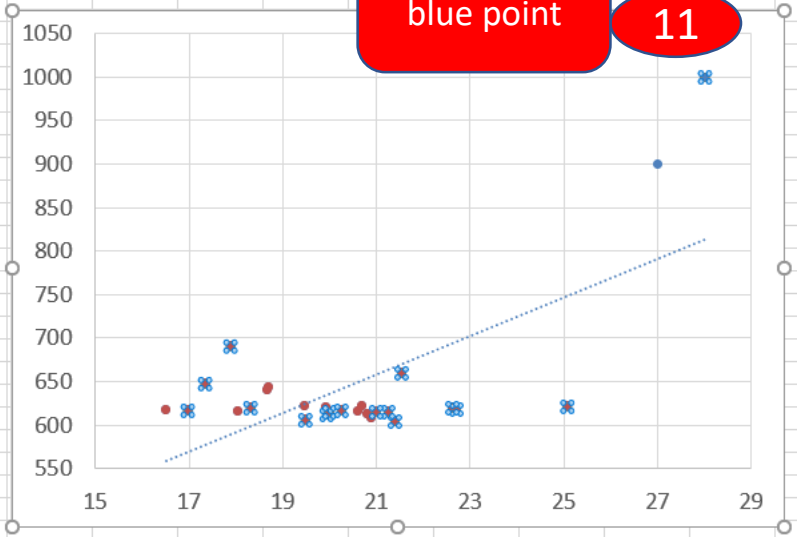


Chart Elements
- Axes
- Axis Titles
- Chart Title
- Data Labels
- Error Bars
- Gridlines
- Legend
- Trendline

12

caschool | Figure 4.2 | linear_model | Table4.1 | Prediction | outliers

Ready

caschool - Excel — Pravesh Tamang

Chart Tools

File Home Insert Page Layout Formulas Data Review View Developer Help Foxit Reader PDF Design Format Tell me what you want to do Share

Add Chart Element | Quick Layout | Change Colors | Chart Layouts | Chart Styles | Switch Row/ Column | Select Data | Data | Change Chart Type | Type | Move Chart | Location

Chart 2

=SERIES("without_outlier",outliers!$A$2:$A$26,outliers!$B$2:$B$26,2)

	A	B
1	str	testscr
2	17.88991	690.8
3	21.52466	661.2
4	18.69723	643.6
5	17.35714	647.7
6	18.67133	640.85
7	21.40625	605.55
8	19.5	606.75
9	20.89412	609
10	19.94737	612.5
11	20.80556	612.65
12	21.23809	615.75
13	21	616.3
14	20.6	616.3
15	20.00822	616.3
16	18.02778	616.45
17	20.25196	617.35
18	16.97787	618.05
19	16.5098	618.3
20	22.70402	619.8
21	19.91111	620.3
22	18.33333	620.5
23	22.61905	621.4
24	19.44828	621.75
25	25.05263	622.05
26	20.67544	622.6
27	28	1000
28	27	900

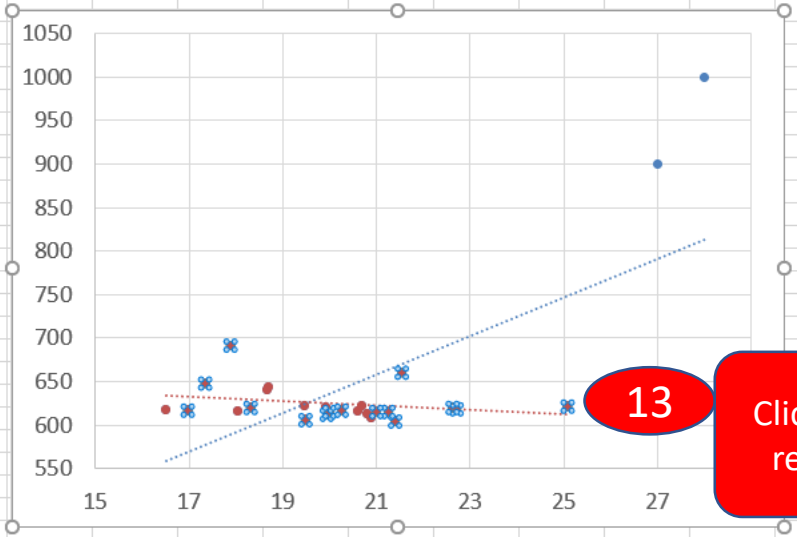


Chart Elements
- [x] Axes
- [ ] Axis Titles
- [ ] Chart Title
- [ ] Data Labels
- [ ] Error Bars
- [x] Gridlines
- [ ] Legend
- [x] Trendline

14

13

Click on the red point

caschool | Figure 4.2 | linear_model | Table4.1 | Prediction | outliers

Ready

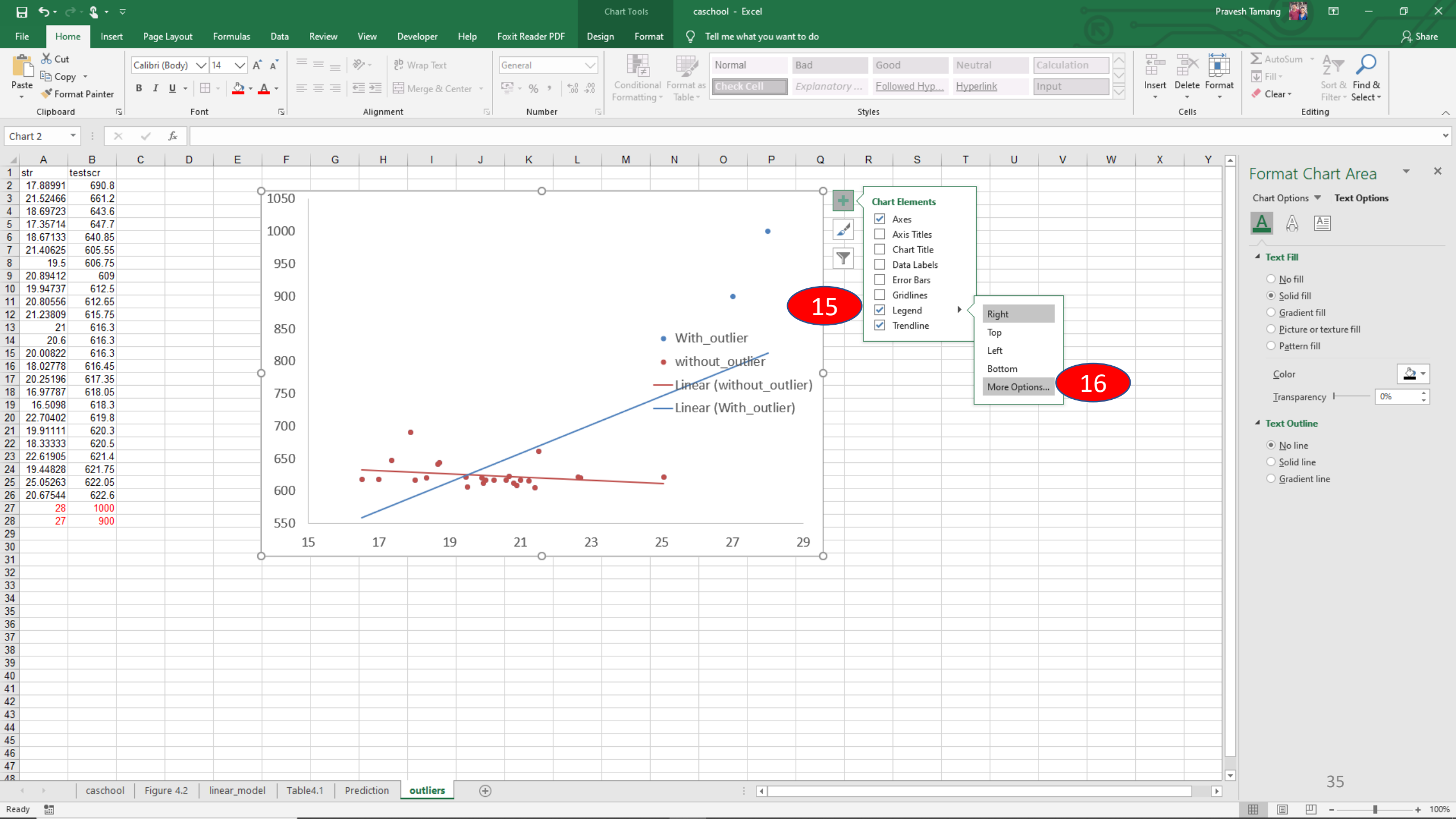

caschool - Excel
Chart Tools
Pravesh Tamang
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Developer
Help
Foxit Reader PDF
Design
Format
Tell me what you want to do
Share
Paste
Cut
Copy
Format Painter
Clipboard
Calibri (Body)
14
Font
Wrap Text
Merge & Center
Alignment
General
Number
Conditional Formatting
Format as Table
Normal
Bad
Good
Neutral
Calculation
Check Cell
Explanatory ...
Followed Hyp...
Hyperlink
Input
Styles
Insert
Delete
Format
Cells
AutoSum
Fill
Clear
Sort & Filter
Find & Select
Editing
Chart 2
str
testscr
17.88991 690.8
21.52466 661.2
18.69723 643.6
17.35714 647.7
18.67133 640.85
21.40625 605.55
19.5 606.75
20.89412 609
19.94737 612.5
20.80556 612.65
21.23809 615.75
21 616.3
20.6 616.3
20.00822 616.3
18.02778 616.45
20.25196 617.35
16.97787 618.05
16.5098 618.3
22.70402 619.8
19.91111 620.3
18.33333 620.5
22.61905 621.4
19.44828 621.75
25.05263 622.05
20.67544 622.6
28 1000
27 900
1050
1000
950
900
850
800
750
700
650
600
550
15
17
19
21
23
25
27
29
With_outlier
without_outlier
Linear (without_outlier)
Linear (With_outlier)
Chart Elements
Axes
Axis Titles
Chart Title
Data Labels
Error Bars
Gridlines
Legend
Trendline
Right
Top
Left
Bottom
More Options...
15
16
Format Chart Area
Chart Options
Text Options
Text Fill
No fill
Solid fill
Gradient fill
Picture or texture fill
Pattern fill
Color
Transparency
0%
Text Outline
No line
Solid line
Gradient line
caschool
Figure 4.2
linear_model
Table4.1
Prediction
outliers
35
Ready
100%

Chart Tools | caschool - Excel | Pravesh Tamang

File | Home | Insert | Page Layout | Formulas | Data | Review | View | Developer | Help | Foxit Reader PDF | Design | Format | Tell me what you want to do | Share

Paste | Cut | Copy | Format Painter | Clipboard

Calibri (Body) | 14 | Font

Wrap Text | Merge & Center | Alignment

General | Number

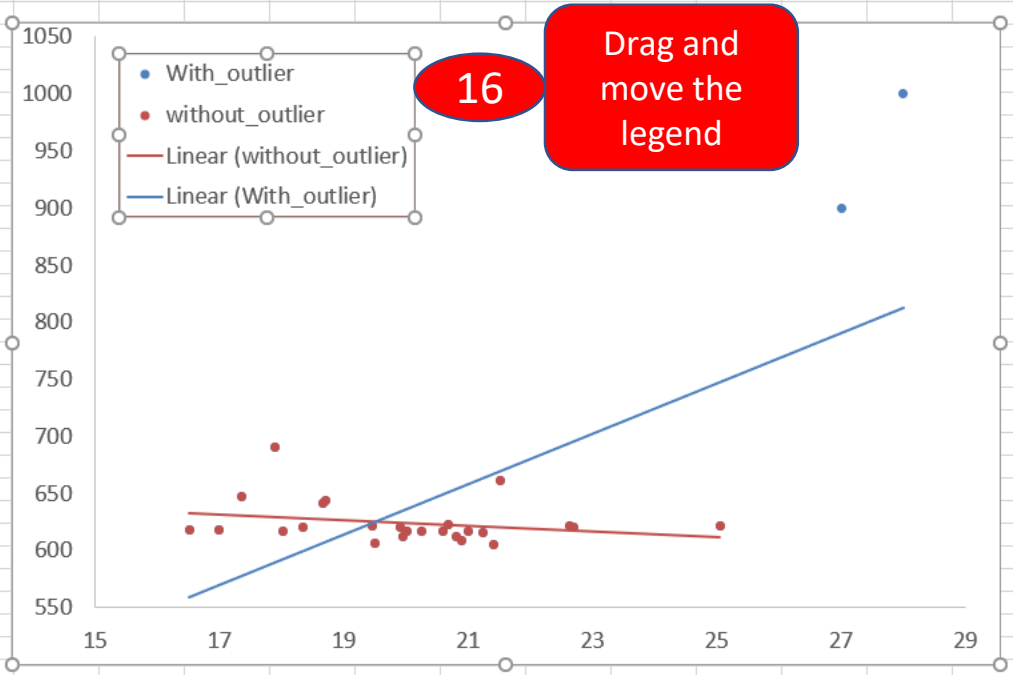
Conditional Formatting | Format as Table | Normal | Bad | Good | Neutral | Calculation | Check Cell | Explanatory ... | Followed Hyp... | Hyperlink | Input | Styles

Insert | Delete | Format | Cells

AutoSum | Fill | Clear | Sort & Filter | Find & Select | Editing

Chart 2

	A	B
1	str	testscr
2	17.88991	690.8
3	21.52466	661.2
4	18.69723	643.6
5	17.35714	647.7
6	18.67133	640.85
7	21.40625	605.55
8	19.5	606.75
9	20.89412	609
10	19.94737	612.5
11	20.80556	612.65
12	21.23809	615.75
13	21	616.3
14	20.6	616.3
15	20.00822	616.3
16	18.02778	616.45
17	20.25196	617.35
18	16.97787	618.05
19	16.5098	618.3
20	22.70402	619.8
21	19.91111	620.3
22	18.33333	620.5
23	22.61905	621.4
24	19.44828	621.75
25	25.05263	622.05
26	20.67544	622.6
27	28	1000
28	27	900



16

Drag and move the legend

Format Legend

Legend Options | Text Options

Fill

- No fill
- Solid fill
- Gradient fill
- Picture or texture fill
- Pattern fill
- Automatic

Color

Border

- No line
- Solid line
- Gradient line
- Automatic

17

Color

Transparency 0%

Width 0.75 pt

Compound type

Dash type

Cap type Flat

Join type Round

Begin Arrow type

Begin Arrow size

End Arrow type

End Arrow size

caschool | Figure 4.2 | linear_model | Table4.1 | Prediction | outliers

Ready

Exercise

E4.1 You will find the data and description file here. Use the data **Growth.xlsx**.

a) Construct a scatterplot of average annual growth rate *(Growth)* on the average trade share *(TradeShare)*. Does there appear to be a relationship between the variables?
b) One country, Malta, has a trade share much larger than the other countries. Find Malta on the scatterplot. Does Malta look like an outlier?
c) Using all observations, run a regression of *Growth* on *TradeShare*. What is the estimated slope? What is the estimated intercept? Use the regression to predict the growth rate for a country with a trade share of 0.5 and with a trade share equal to 1.0.
d) Estimate the same regression, excluding the data from Malta. Answer the same questions in (c).
e) Plot the estimated regression functions from (c) and (d). Using the scatterplot in (a), explain why the regression function that includes Malta is steeper than the regression function that excludes Malta.
f) Where is Malta? Why is the Malta trade share so large? Should Malta be included or excluded from the analysis?

E4.2 You will find the data and description file here. Use the data **Earnings_and_Height.xlsx**.

a) What is the median value of height in the sample?

b) (i). Estimate average earnings for workers whose height is at most 67 inches.

(ii). Estimate average earnings for workers whose height is greater than 67 inches.

(iii). On average, do taller workers earn more than shorter workers? How much more? What is a 95% confidence interval for the difference in average earnings?

c) Construct a scatterplot of annual earnings (*Earnings*) on height (*Height*). Notice that the points on the plot fall along horizontal lines. (There are only 23 distinct values of Earnings). Why? (Hint: Carefully read the detailed data description.)

d) Run a regression of Earnings on Height.

(i). What is the estimated slope?

(ii). Use the estimated regression to predict earnings for a worker who is 67 inches tall, for a worker who is 70 inches tall, and for a worker who is 65 inches tall.

e) Suppose height were measured in centimeters instead of inches. Answer the following questions about the Earnings on Height (in cm) regression.

(i). What is the estimated slope of the regression?

(ii). What is the estimated intercept?

(iii). What is the R2?

(iv). What is the standard error of the regression?

(f). Run a regression of Earnings on Height, using data for female workers only.
(i). What is the estimated slope?
(ii). A randomly selected woman is 1 inch taller than the average woman in the sample. Would you predict her earnings to be higher or lower than the average earnings for women in the sample? By how much?
(g). Repeat (f) for male workers.
(h). Do you think that height is uncorrelated with other factors that cause earning? That is, do you think that the regression error term, say u_i, has a conditional mean of zero, given Height (X_i)?