Solayman Hossain Emon

ID: 80744292

In [232…

```
using CSV, DataFrames, GLM, RegressionTables, Gadfly, Plots
```

In [233…

```
# Enable printing of 1000 columns
ENV["COLUMNS"] = 1000
```

Out[233…

```
1000
```

Reading the Data

In [234…

```
fb_data = CSV.read("dataset_Facebook.csv", DataFrame)
```

Out[234… 500 rows × 19 columns

	Page total likes	Type	Category	Post Month	Post Weekday	Post Hour	Paid	Lifetime Post Total Reach	Lifetime Post Total Impressions	Lifetime Engaged Users	Lifetime Post Consumers	Lifetime Post Consumptions	Lifetime Post Impressions by people who have liked your Page
	Int64	String7	Int64	Int64	Int64	Int64	Int64?	Int64	Int64	Int64	Int64	Int64	Int64
1	139441	Photo	2	12	4	3	0	2752	5091	178	109	159	3078
2	139441	Status	2	12	3	10	0	10460	19057	1457	1361	1674	11710
3	139441	Photo	3	12	3	3	0	2413	4373	177	113	154	2812

4	139441	Photo	2	12	2	10	1	50128	87991	2211	790	1119	61027	3
5	139441	Photo	2	12	2	3	0	7244	13594	671	410	580	6228	
6	139441	Status	2	12	1	9	0	10472	20849	1191	1073	1389	16034	
7	139441	Photo	3	12	1	3	1	11692	19479	481	265	364	15432	
8	139441	Photo	3	12	7	9	1	13720	24137	537	232	305	19728	1
9	139441	Status	2	12	7	3	0	11844	22538	1530	1407	1692	15220	
10	139441	Photo	3	12	6	10	0	4694	8668	280	183	250	4309	
11	139441	Status	2	12	5	10	0	21744	42334	4258	4100	4540	37849	1
12	139441	Photo	2	12	5	10	0	3112	5590	208	127	145	3887	
13	139441	Photo	2	12	5	10	0	2847	5133	193	115	133	3779	
14	139441	Photo	2	12	5	3	0	2549	4896	249	134	168	3631	
15	138414	Photo	2	12	4	5	1	22784	39941	887	337	417	34415	1
16	138414	Status	2	12	3	10	0	10060	19680	1264	1209	1425	17272	
17	138414	Photo	3	12	3	3	0	1722	2981	163	123	148	1868	
18	138414	Photo	1	12	2	12	1	53264	111785	1706	1103	1655	92512	3
19	138414	Status	3	12	2	3	0	3930	7509	130	86	112	5009	
20	138414	Photo	3	12	1	11	0	1591	2825	121	88	111	2116	
21	138414	Photo	2	12	1	3	0	2848	5066	200	142	184	3561	
22	138414	Photo	1	12	7	10	0	1384	2467	15	15	20	2196	
23	138414	Link	1	12	7	10	0	3454	6853	118	104	130	6282	
24	138414	Photo	3	12	7	3	0	2723	4888	176	118	143	2964	
25	138414	Status	2	12	6	10	0	8488	15294	1341	1270	1489	9684	
26	138458	Status	2	12	6	3	0	8284	15104	1521	1462	1711	10266	
27	138458	Status	2	12	5	11	0	19552	34143	2806	2531	3420	17748	
28	138458	Photo	3	12	5	3	0	2478	4306	212	124	149	2612	
29	138895	Photo	2	12	5	3	0	9560	18264	973	559	885	9217	

30	138895	Video	1	12	4	11	1	36208	61262	1141	1068	1728	30131	1.
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

In [235...

```
println(size(fb_data))
```

```
(500, 19)
```

In [236...

```
names(fb_data)
```

Out[236...

```
19-element Vector{String}:
 "Page total likes"
 "Type"
 "Category"
 "Post Month"
 "Post Weekday"
 "Post Hour"
 "Paid"
 "Lifetime Post Total Reach"
 "Lifetime Post Total Impressions"
 "Lifetime Engaged Users"
 "Lifetime Post Consumers"
 "Lifetime Post Consumptions"
 "Lifetime Post Impressions by people who have liked your Page"
 "Lifetime Post reach by people who like your Page"
 "Lifetime People who have liked your Page and engaged with your post"
 "comment"
 "like"
 "share"
 "Total Interactions"
```

In [237...

```
describe(fb_data)
```

Out[237...

19 rows × 7 columns

	variable	mean	min	median	max	nmissing	eltype
	Symbol	Union...	Any	Union...	Any	Int64	Type
1	Page total likes	1.23194e5	81370	129600.0	139441	0	Int64

	variable	mean	min	median	max	nmissing	eltype
	Symbol	Union...	Any	Union...	Any	Int64	Type
2	Type		Link		Video	0	String7
3	Category	1.88	1	2.0	3	0	Int64
4	Post Month	7.038	1	7.0	12	0	Int64
5	Post Weekday	4.15	1	4.0	7	0	Int64
6	Post Hour	7.84	1	9.0	23	0	Int64
7	Paid	0.278557	0	0.0	1	1	Union{Missing, Int64}
8	Lifetime Post Total Reach	13903.4	238	5281.0	180480	0	Int64
9	Lifetime Post Total Impressions	29585.9	570	9051.0	1110282	0	Int64
10	Lifetime Engaged Users	920.344	9	625.5	11452	0	Int64
11	Lifetime Post Consumers	798.772	9	551.5	11328	0	Int64
12	Lifetime Post Consumptions	1415.13	9	851.0	19779	0	Int64
13	Lifetime Post Impressions by people who have liked your Page	16766.4	567	6255.5	1107833	0	Int64
14	Lifetime Post reach by people who like your Page	6585.49	236	3417.0	51456	0	Int64
15	Lifetime People who have liked your Page and engaged with your post	609.986	9	412.0	4376	0	Int64
16	comment	7.482	0	3.0	372	0	Int64
17	like	177.946	0	101.0	5172	1	Union{Missing, Int64}
18	share	27.2661	0	19.0	790	4	Union{Missing, Int64}
19	Total Interactions	212.12	0	123.5	6334	0	Int64

Handling the Columns with Space

In [238…

```
rename!(fb_data, Dict(:"Page total likes" => :"Page_total_likes",
                      :"Post Month" => :"Post_Month",
                      :"Post Hour" => :"Post_Hour",
                      :"Post Weekday" => :"Post_Weekday"));
```

In [239…

```
names(fb_data)
```

Out[239…

```
19-element Vector{String}:
 "Page_total_likes"
 "Type"
 "Category"
 "Post_Month"
 "Post_Weekday"
 "Post_Hour"
 "Paid"
 "Lifetime Post Total Reach"
 "Lifetime Post Total Impressions"
 "Lifetime Engaged Users"
 "Lifetime Post Consumers"
 "Lifetime Post Consumptions"
 "Lifetime Post Impressions by people who have liked your Page"
 "Lifetime Post reach by people who like your Page"
 "Lifetime People who have liked your Page and engaged with your post"
 "comment"
 "like"
 "share"
 "Total Interactions"
```

In [240…

```
dropmissing!(fb_data, :"like");
dropmissing!(fb_data, :"Paid");
dropmissing!(fb_data, :"share");
describe(fb_data)
```

Out[240… 19 rows × 7 columns

	variable	mean	min	median	max	nmissing	eltype
	Symbol	Union…	Any	Union…	Any	Int64	DataType
1	Page_total_likes	1.23173e5	81370	129600.0	139441	0	Int64
2	Type		Link		Video	0	String7
3	Category	1.88687	1	2.0	3	0	Int64
4	Post_Month	7.02828	1	7.0	12	0	Int64
5	Post_Weekday	4.13333	1	4.0	7	0	Int64
6	Post_Hour	7.84444	1	9.0	23	0	Int64

	variable	mean	min	median	max	nmissing	eltype
	Symbol	Union...	Any	Union...	Any	Int64	DataType
7	Paid	0.280808	0	0.0	1	0	Int64
8	Lifetime Post Total Reach	14028.1	238	5290.0	180480	0	Int64
9	Lifetime Post Total Impressions	29857.0	570	9084.0	1110282	0	Int64
10	Lifetime Engaged Users	926.83	9	630.0	11452	0	Int64
11	Lifetime Post Consumers	804.156	9	555.0	11328	0	Int64
12	Lifetime Post Consumptions	1425.92	9	861.0	19779	0	Int64
13	Lifetime Post Impressions by people who have liked your Page	16916.3	567	6282.0	1107833	0	Int64
14	Lifetime Post reach by people who like your Page	6641.36	236	3478.0	51456	0	Int64
15	Lifetime People who have liked your Page and engaged with your post	614.135	9	416.0	4376	0	Int64
16	comment	7.55758	0	3.0	372	0	Int64
17	like	179.145	0	101.0	5172	0	Int64
18	share	27.2646	0	19.0	790	0	Int64
19	Total Interactions	213.968	0	125.0	6334	0	Int64

1. Multiple Regression Model

(a) Fit the Model

In [241…

```
m1 = @formula(like~Category+Page_total_likes+Type+Post_Month+Post_Hour+Post_Weekday+Paid+comment)
```

Out[241…

```
FormulaTerm
Response:
  like(unknown)
Predictors:
  Category(unknown)
  Page_total_likes(unknown)
  Type(unknown)
  Post_Month(unknown)
```

```
Post_Hour(unknown)
Post_Weekday(unknown)
Paid(unknown)
comment(unknown)
```

In [242…

```
model1 = lm(m1, fb_data)
```

Out[242…

```
StatsModels.TableRegressionModel{LinearModel{GLM.LmResp{Vector{Float64}}, GLM.DensePredChol{Float64, CholeskyPivoted{Float64, Matrix{Float64}}}}, Matrix{Float64}}

like ~ 1 + Category + Page_total_likes + Type + Post_Month + Post_Hour + Post_Weekday + Paid + comment

Coefficients:
──────────────────────────────────────────────────────────────────────────────────────────
                         Coef.     Std. Error       t  Pr(>|t|)       Lower 95%     Upper 95%
──────────────────────────────────────────────────────────────────────────────────────────
(Intercept)        -76.0596       134.32          -0.57    0.5715  -339.982       187.863
Category            40.9061         9.67147        4.23    <1e-04    21.9028       59.9093
Page_total_likes     0.000477266    0.00146798     0.33    0.7452    -0.00240714    0.00336168
Type: Photo         15.2491        39.8225         0.38    0.7019   -62.9973       93.4955
Type: Status       -19.0594        47.4556        -0.40    0.6881  -112.304        74.185
Type: Video         19.5248        76.8518         0.25    0.7996  -131.48        170.529
Post_Month           1.70077        7.25882        0.23    0.8148   -12.5619       15.9635
Post_Hour           -0.406625       1.86435       -0.22    0.8274    -4.06984       3.25659
Post_Weekday        -1.58256        3.89916       -0.41    0.6850    -9.24394       6.07881
Paid                35.1066        17.6526         1.99    0.0473     0.421469     69.7917
comment             12.6549         0.371847      34.03    <1e-99    11.9243       13.3856
──────────────────────────────────────────────────────────────────────────────────────────
```

R-Squared Values

In [243…

```
# R Square value of the model
r2(model1)
```

Out[243…

```
0.717372980012585
```

Coefficient Significance

In [244…

```
coef(model1)
```

Out[244…
```
11-element Vector{Float64}:
 -76.05956033276802
  40.90605304627139
   0.00047726636210155047
  15.249068574473092
 -19.05938067480692
  19.52481991626934
   1.7007699271319174
  -0.4066248513377943
  -1.5825611175721401
  35.10660018204332
  12.654937461963543
```

Refit with Logarithm Transformation

In [245…
```
m1_log = @formula(log1p(like)~Category+log1p(Page_total_likes)+Type+Post_Month+Post_Hour+Post_Weekday+Paid)
model1_log = lm(mm, fb_data)
```

Out[245…
```
StatsModels.TableRegressionModel{LinearModel{GLM.LmResp{Vector{Float64}}, GLM.DensePredChol{Float64, CholeskyPivoted{Float64, Matrix{Float64}}}}, Matrix{Float64}}

:(log1p(like)) ~ 1 + Category + :(log1p(Page_total_likes)) + Type + Post_Month + Post_Hour + Post_Weekday + Paid

Coefficients:
```

	Coef.	Std. Error	t	Pr(>|t|)	Lower 95%	Upper 95%
(Intercept)	-9.69856	10.5257	-0.92	0.3573	-30.3801	10.983
Category	0.40557	0.0610486	6.64	<1e-10	0.285618	0.525523
log1p(Page_total_likes)	1.15958	0.924691	1.25	0.2104	-0.657319	2.97647
Type: Photo	0.444327	0.251196	1.77	0.0775	-0.0492398	0.937893
Type: Status	0.709225	0.299309	2.37	0.0182	0.121122	1.29733
Type: Video	1.3856	0.48481	2.86	0.0044	0.433013	2.33819
Post_Month	-0.0377243	0.0403821	-0.93	0.3507	-0.11707	0.0416211
Post_Hour	-0.0155737	0.0117812	-1.32	0.1868	-0.0387222	0.00757471
Post_Weekday	-0.0559603	0.0245616	-2.28	0.0231	-0.104221	-0.00770015
Paid	0.285997	0.111235	2.57	0.0104	0.0674344	0.504559

In [246…
```
# R Square value of the model
r2(model1_log)
```

```
0.1479247827797423
```

Out[246…

In [247…

```
coef(model1_log)
```

Out[247…

```
10-element Vector{Float64}:
 -9.698556404446071
  0.40557034980235973
  1.1595758609956286
  0.44432673727588945
  0.7092246134950749
  1.3856007166483466
 -0.03772427706108226
 -0.015573736135879092
 -0.055960342598400256
  0.2859968739277099
```

Models Dignostics

In [248…

```
regtable(model1, model1_log)
```

```
-------------------------------------------------
                                like   log1p(like)
                           ---------   -----------
                                 (1)           (2)
-------------------------------------------------
(Intercept)                  -76.060        -9.699
                           (134.320)      (10.526)
Category                   40.906***      0.406***
                             (9.671)       (0.061)
Page_total_likes               0.000
                             (0.001)
Type: Photo                   15.249         0.444
                            (39.823)       (0.251)
Type: Status                 -19.059        0.709*
                            (47.456)       (0.299)
Type: Video                   19.525       1.386**
                            (76.852)       (0.485)
Post_Month                     1.701        -0.038
                             (7.259)       (0.040)
Post_Hour                     -0.407        -0.016
                             (1.864)       (0.012)
Post_Weekday                  -1.583       -0.056*
```

```
                          (3.899)        (0.025)
Paid                      35.107*        0.286*
                         (17.653)        (0.111)
comment                  12.655***
                          (0.372)
log1p(Page_total_likes)                   1.160
                                         (0.925)
-------------------------------------------------
Estimator                     OLS            OLS
-------------------------------------------------
N                             495            495
R2                          0.717          0.148
-------------------------------------------------
```

(C) Assumptions

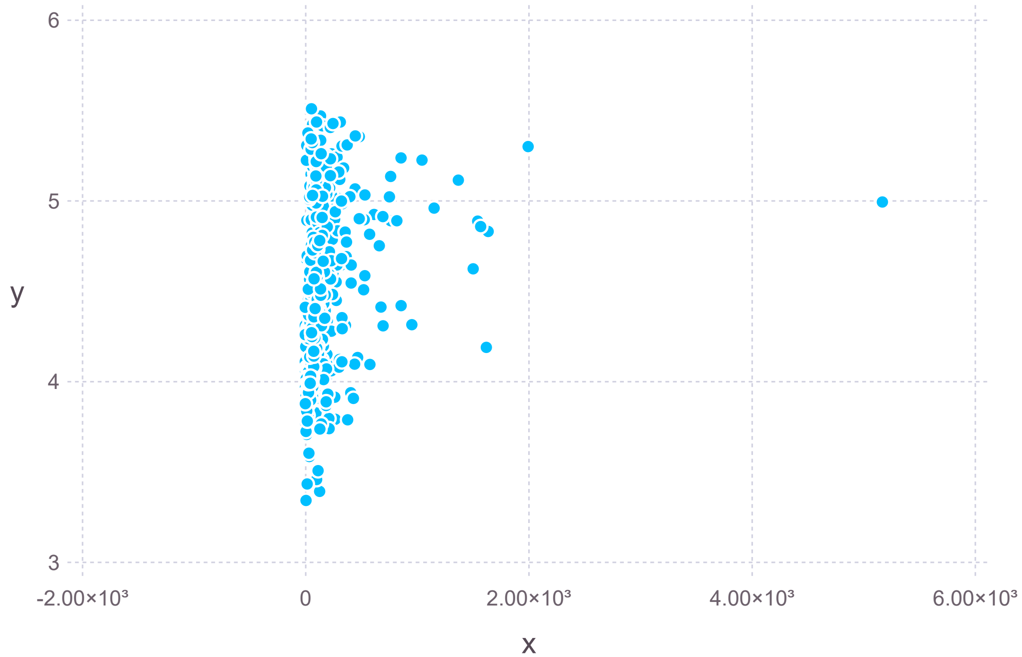
In [249…

```
# predicted values
pred1 = predict(model1_log)
# residuals
red1 = fb_data[!, :like] - pred1;
```

In [250…

```
# homoscedasticity Assumption
plot(x=red1,y=pred1, Geom.point)
```

Out[250…

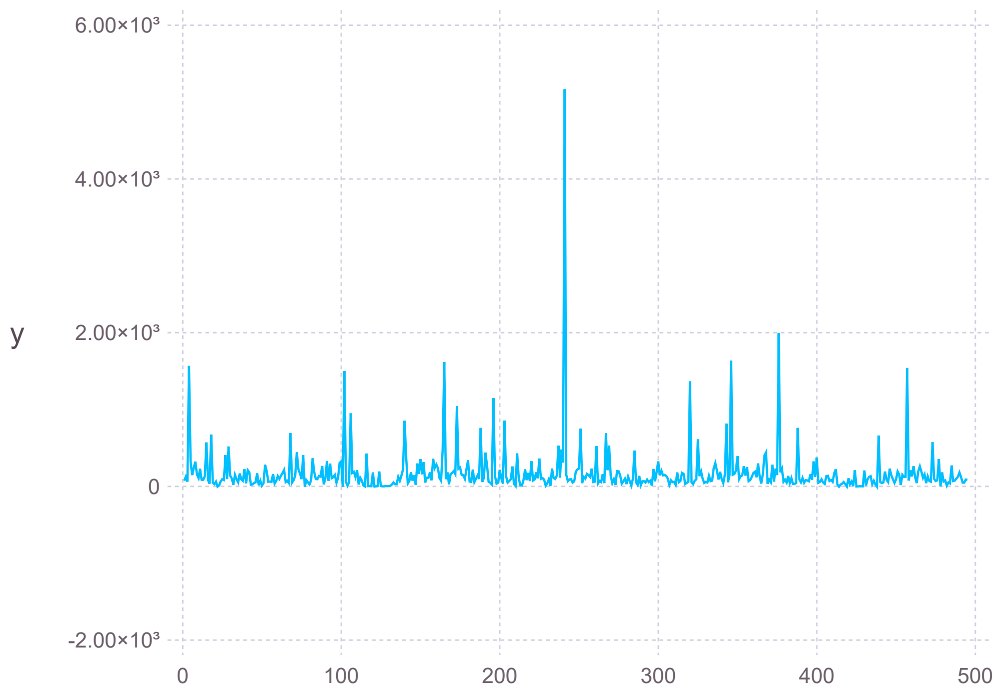


Here the variance is not same in different X value, so homoscedasticity assumption is not OK

In [251…

```
# Independence Assumption
plot(y=red1, Geom.line)
```

Out[251…

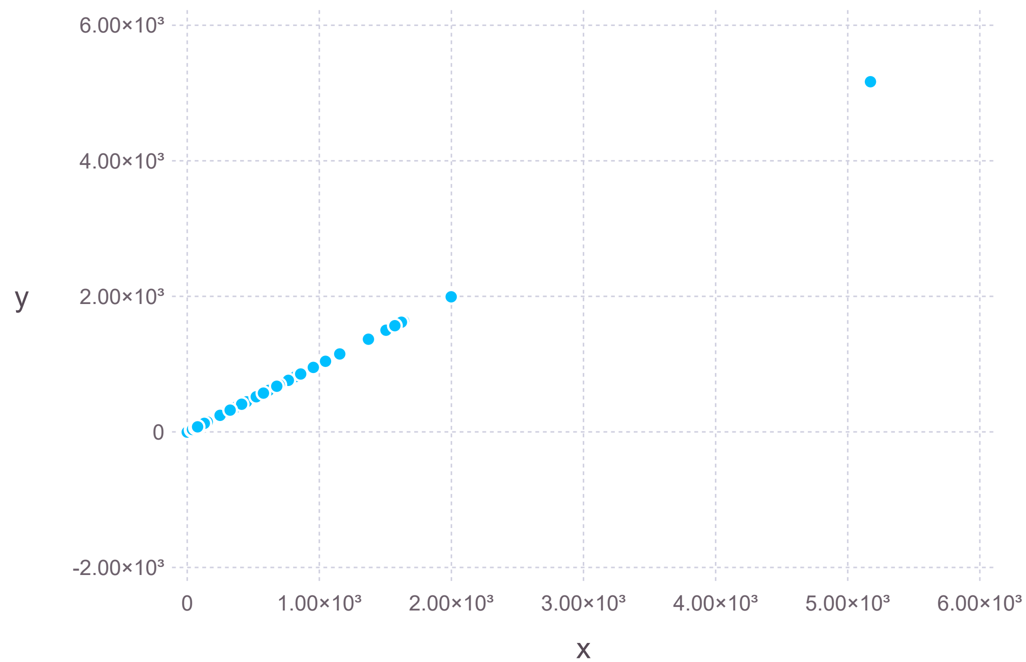


All the observations look random, So the independent assumption hold

In [252…

```
# Exogeneity assumption
plot(x = fb_data[!, :like], y=red1)
```

Out[252…



We observed some pattern over the plot, so the Epsilon captured some information with was left out previously. It violating the assumption & might missed some idependent varaiable

Adding Some more varaiable for Observing the Diagnostic Changes in the Model

In [253…

```
rename!(fb_data, Dict(:"Total Interactions" => :"Total_Interactions"));
```

In [254…

```
m1_test = @formula(log1p(like)~Category+log1p(Page_total_likes)+Type+Post_Month+Post_Hour+Post_Weekday+Paid+Total_Interac
model1_test = lm(m1_test, fb_data)
```

Out[254…

```
StatsModels.TableRegressionModel{LinearModel{GLM.LmResp{Vector{Float64}}, GLM.DensePredChol{Float64, CholeskyPivoted{Floa
t64, Matrix{Float64}}}}, Matrix{Float64}}

:(log1p(like)) ~ 1 + Category + :(log1p(Page_total_likes)) + Type + Post_Month + Post_Hour + Post_Weekday + Paid + Total_
Interactions

Coefficients:
```

	Coef.	Std. Error	t	Pr(>|t|)	Lower 95%	Upper 95%
(Intercept)	-3.25106	8.61178	-0.38	0.7060	-20.1722	13.67
Category	0.315722	0.0502231	6.29	<1e-09	0.217039	0.414404
log1p(Page_total_likes)	0.587934	0.75657	0.78	0.4375	-0.898632	2.0745
Type: Photo	0.329696	0.205415	1.61	0.1091	-0.0739186	0.733311
Type: Status	0.601503	0.2447	2.46	0.0143	0.120698	1.08231
Type: Video	1.12752	0.396544	2.84	0.0047	0.348364	1.90669
Post_Month	-0.0215118	0.0330175	-0.65	0.5150	-0.0863873	0.0433636
Post_Hour	-0.0146333	0.00962802	-1.52	0.1292	-0.0335512	0.0042846
Post_Weekday	-0.0346999	0.0201187	-1.72	0.0852	-0.0742306	0.00483083
Paid	0.138856	0.0913942	1.52	0.1293	-0.040722	0.318435
Total_Interactions	0.00168251	0.000108109	15.56	<1e-43	0.00147009	0.00189493

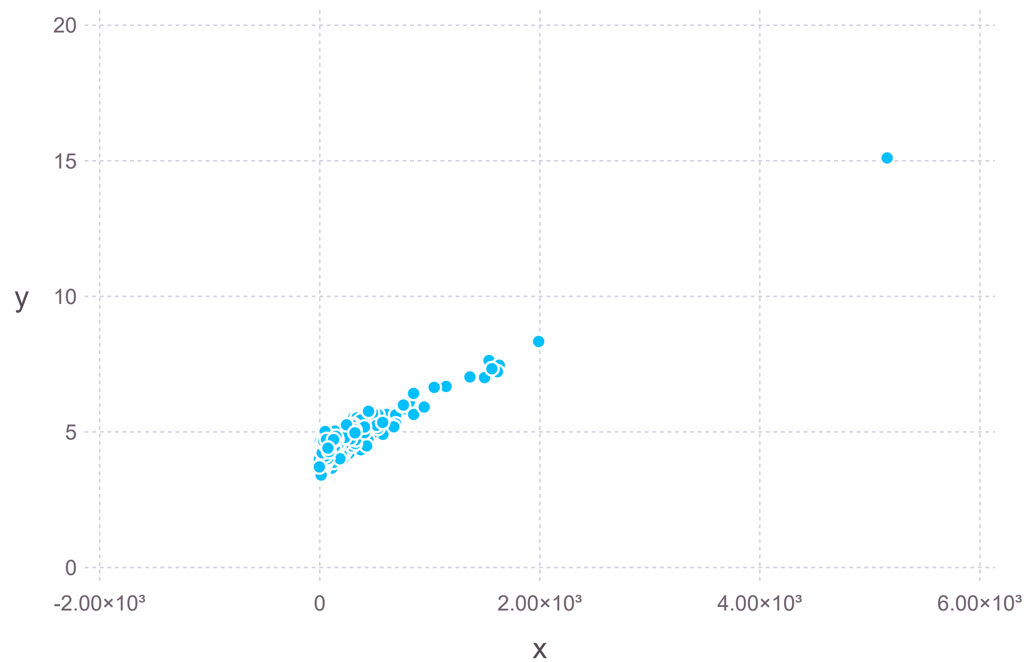
In [255…

```
# predicted values
pred1_test = predict(model1_test)
# residuals
red1_test = fb_data[!, :like] - pred1_test;
```

In [256…

```
# homoscedasticity Assumption
plot(x=red1_test, y=pred1_test, Geom.point)
```

Out[256…

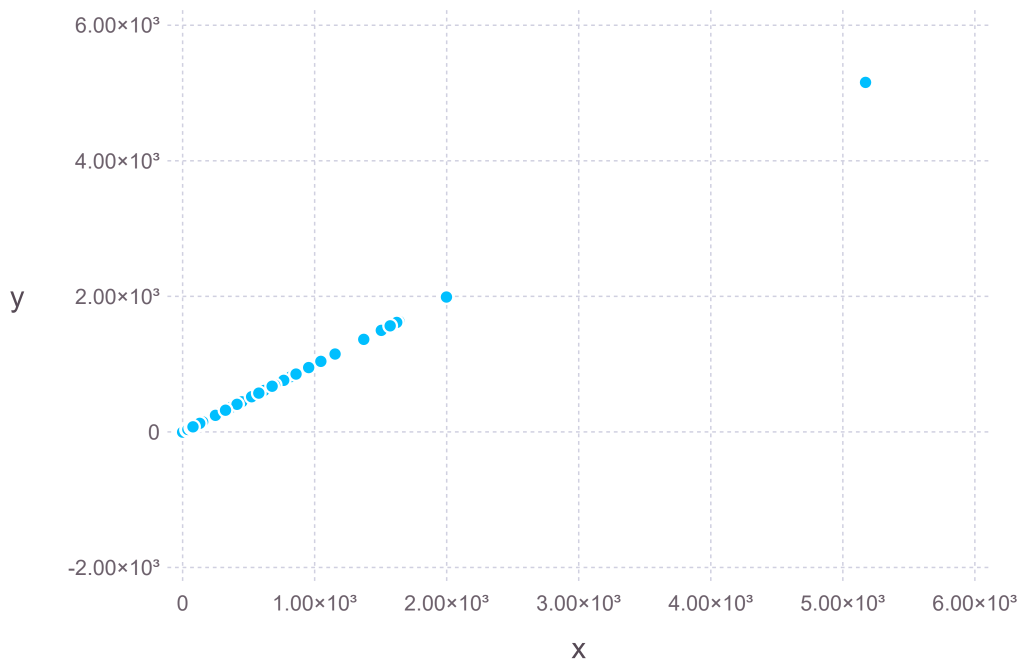


We can observed that after adding more variables (not all) the homoscedasticity assumption behavior is varying from the previous

In [257…

```
# Exogeneity assumption
plot(x = fb_data[!, :like], y=red1_test)
```

Out[257…



The Exogeneity assumption is not changed from the previous

In [258…

```
regtable(model1, model1_log, model1_test)
```

```
---------------------------------------------------------
                              like       log1p(like)
                         ---------   -------------------
                               (1)         (2)        (3)
---------------------------------------------------------
(Intercept)                -76.060      -9.699     -3.251
                         (134.320)    (10.526)    (8.612)
Category                 40.906***    0.406***   0.316***
                           (9.671)     (0.061)    (0.050)
Page_total_likes             0.000
                           (0.001)
Type: Photo                 15.249       0.444      0.330
                          (39.823)     (0.251)    (0.205)
Type: Status               -19.059      0.709*     0.602*
                          (47.456)     (0.299)    (0.245)
```

```
Type: Video                  19.525     1.386**    1.128**
                            (76.852)   (0.485)    (0.397)
Post_Month                    1.701     -0.038     -0.022
                             (7.259)   (0.040)    (0.033)
Post_Hour                    -0.407     -0.016     -0.015
                             (1.864)   (0.012)    (0.010)
Post_Weekday                 -1.583     -0.056*    -0.035
                             (3.899)   (0.025)    (0.020)
Paid                         35.107*     0.286*     0.139
                            (17.653)   (0.111)    (0.091)
comment                     12.655***
                             (0.372)
log1p(Page_total_likes)                 1.160      0.588
                                       (0.925)    (0.757)
Total_Interactions                                 0.002***
                                                  (0.000)
-------------------------------------------------------------
Estimator                       OLS        OLS        OLS
-------------------------------------------------------------
N                               495        495        495
R2                            0.717      0.148      0.432
-------------------------------------------------------------
```

After doing Logarithm transformation and adding some Extra variable from the dataset the performanace of the Multiple Linear Regression is increasing and some assumptions are really closed to be hold. So, we can assume that doing some appropriate transformation and evaluating assumption after adding variables will be the possible remedy to handle this.

(b) Dropping Some Variables and Observe the Result

In [259…

```
m1_drop = @formula(log1p(like)~Category+Type+Paid+Total_Interactions)
model1_drop = lm(m1_drop, fb_data)
```

Out[259…

```
StatsModels.TableRegressionModel{LinearModel{GLM.LmResp{Vector{Float64}}, GLM.DensePredChol{Float64, CholeskyPivoted{Floa
t64, Matrix{Float64}}}}, Matrix{Float64}}

:(log1p(like)) ~ 1 + Category + Type + Paid + Total_Interactions
```

Coefficients:

	Coef.	Std. Error	t	Pr(>|t|)	Lower 95%	Upper 95%
(Intercept)	3.20202	0.201682	15.88	<1e-45	2.80574	3.59829
Category	0.329537	0.0492241	6.69	<1e-10	0.23282	0.426254
Type: Photo	0.318409	0.201101	1.58	0.1140	-0.0767211	0.713539
Type: Status	0.587141	0.238663	2.46	0.0142	0.118207	1.05607
Type: Video	1.13495	0.392664	2.89	0.0040	0.363431	1.90647
Paid	0.151624	0.0910276	1.67	0.0964	-0.0272303	0.330479
Total_Interactions	0.00170306	0.000107904	15.78	<1e-44	0.00149105	0.00191508

In [260…

```
regtable(model1, model1_log, model1_test, model1_drop)
```

	like	log1p(like)		
	(1)	(2)	(3)	(4)
(Intercept)	-76.060	-9.699	-3.251	3.202***
	(134.320)	(10.526)	(8.612)	(0.202)
Category	40.906***	0.406***	0.316***	0.330***
	(9.671)	(0.061)	(0.050)	(0.049)
Page_total_likes	0.000			
	(0.001)			
Type: Photo	15.249	0.444	0.330	0.318
	(39.823)	(0.251)	(0.205)	(0.201)
Type: Status	-19.059	0.709*	0.602*	0.587*
	(47.456)	(0.299)	(0.245)	(0.239)
Type: Video	19.525	1.386**	1.128**	1.135**
	(76.852)	(0.485)	(0.397)	(0.393)
Post_Month	1.701	-0.038	-0.022	
	(7.259)	(0.040)	(0.033)	
Post_Hour	-0.407	-0.016	-0.015	
	(1.864)	(0.012)	(0.010)	
Post_Weekday	-1.583	-0.056*	-0.035	
	(3.899)	(0.025)	(0.020)	
Paid	35.107*	0.286*	0.139	0.152
	(17.653)	(0.111)	(0.091)	(0.091)
comment	12.655***			
	(0.372)			
log1p(Page_total_likes)		1.160	0.588	

```
                                          (0.925)    (0.757)
Total_Interactions                                   0.002***   0.002***
                                                     (0.000)    (0.000)
-----------------------------------------------------------------------
Estimator                           OLS        OLS        OLS        OLS
-----------------------------------------------------------------------
N                                   495        495        495        495
R2                                0.717      0.148      0.432      0.425
-----------------------------------------------------------------------
```

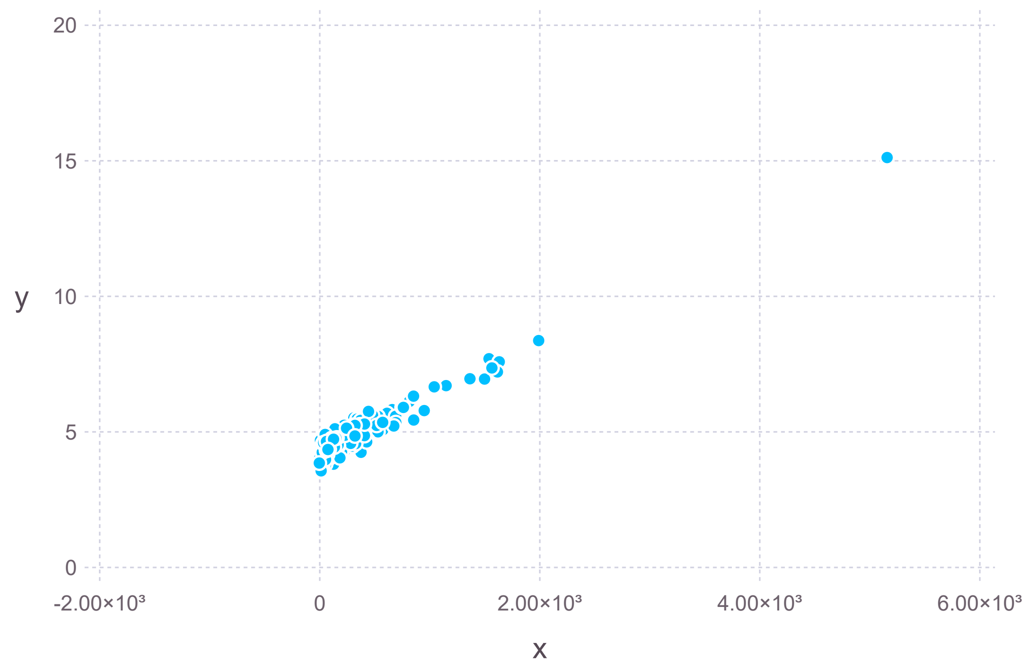
In [261…

```
# predicted values
pred1_drop = predict(model1_drop)
# residuals
red1_drop = fb_data[!, :like] - pred1_drop;
```

In [262…

```
# homoscedasticity Assumption
plot(x=red1_drop, y=pred1_drop, Geom.point)
```

Out[262…



After droppoing some variables from the model, we don't find any significant improvement on the basis of R2 value and Assumptions. So, directly dropping variables might not be a convenient remedy for the regression problem. Otherwise, fixing is more effective way to handle this

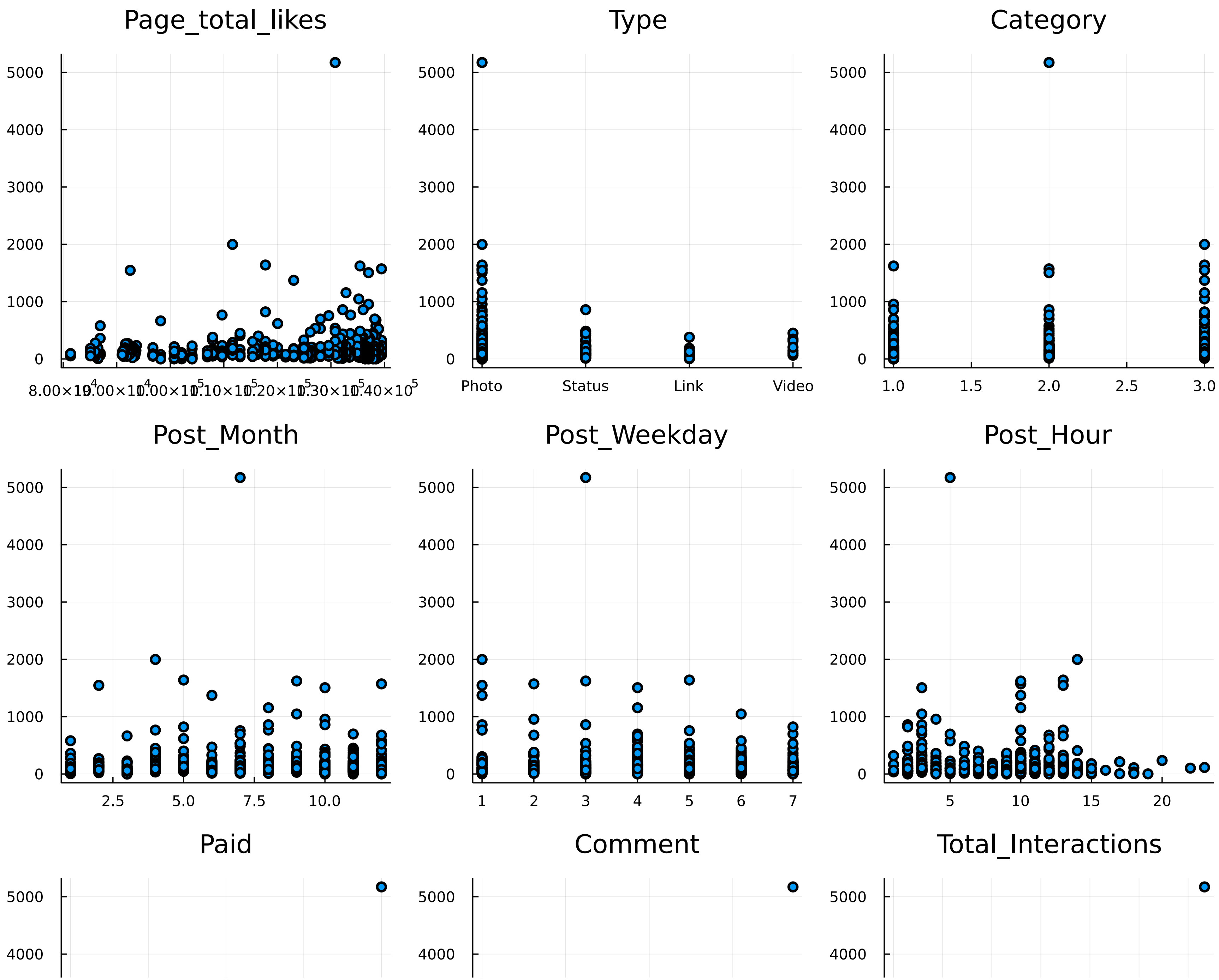
(d) Non-Linear Association

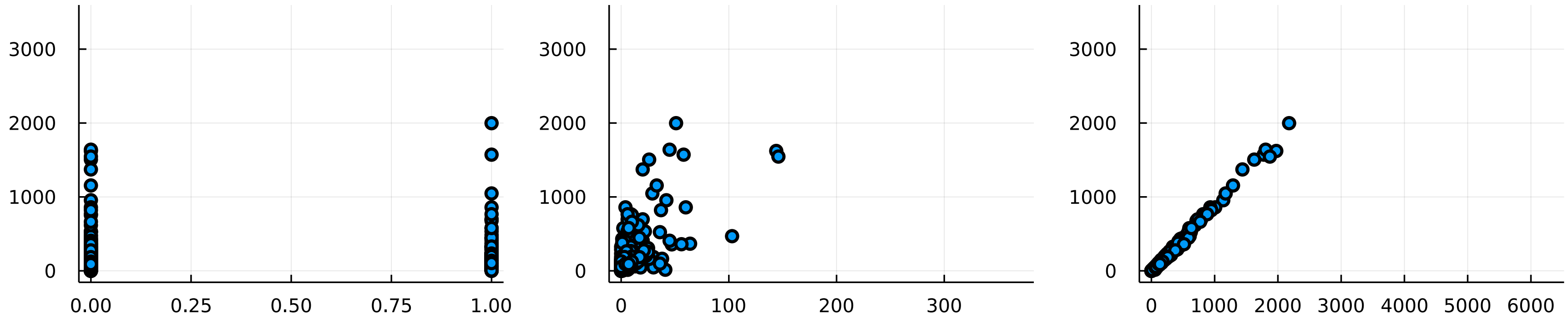
In [263…

```
p1 = scatter(fb_data[!,:Page_total_likes], fb_data[!,:like], title = "Page_total_likes");
p2 = scatter(fb_data[!,:Type], fb_data[!,:like], title = "Type");
p3 = scatter(fb_data[!,:Category], fb_data[!,:like], title = "Category");
p4 = scatter(fb_data[!,:Post_Month], fb_data[!,:like], title = "Post_Month");
p5 = scatter(fb_data[!,:Post_Weekday], fb_data[!,:like], title = "Post_Weekday");
p6 = scatter(fb_data[!,:Post_Hour], fb_data[!,:like], title = "Post_Hour");
p7 = scatter(fb_data[!,:Paid], fb_data[!,:like], title = "Paid");
p8 = scatter(fb_data[!,:comment], fb_data[!,:like], title = "Comment");
p9 = scatter(fb_data[!,:Total_Interactions], fb_data[!,:like], title = "Total_Interactions");
```

```
Plots.plot(p1, p2, p3, p4, p5, p6, p7, p8, p9, layout = (3, 3), legend = false)
plot!(size=(1000, 1000))
```

Out[263…





In [264…

```
names(fb_data)
```

Out[264…

```
19-element Vector{String}:
 "Page_total_likes"
 "Type"
 "Category"
 "Post_Month"
 "Post_Weekday"
 "Post_Hour"
 "Paid"
 "Lifetime Post Total Reach"
 "Lifetime Post Total Impressions"
 "Lifetime Engaged Users"
 "Lifetime Post Consumers"
 "Lifetime Post Consumptions"
 "Lifetime Post Impressions by people who have liked your Page"
 "Lifetime Post reach by people who like your Page"
 "Lifetime People who have liked your Page and engaged with your post"
 "comment"
 "like"
 "share"
 "Total_Interactions"
```

2. Repeat Procedure with Another Outcome Variables

In [265…

```
m2 = @formula(comment~Category+Page_total_likes+Type+Post_Month+Post_Hour+Post_Weekday+Paid)
```

Out[265…

```
FormulaTerm
Response:
  comment(unknown)
```

```
Predictors:
  Category(unknown)
  Page_total_likes(unknown)
  Type(unknown)
  Post_Month(unknown)
  Post_Hour(unknown)
  Post_Weekday(unknown)
  Paid(unknown)
```

In [266…

```
model2 = lm(m2, fb_data)
```

Out[266…

```
StatsModels.TableRegressionModel{LinearModel{GLM.LmResp{Vector{Float64}}, GLM.DensePredChol{Float64, CholeskyPivoted{Float64, Matrix{Float64}}}}, Matrix{Float64}}

comment ~ 1 + Category + Page_total_likes + Type + Post_Month + Post_Hour + Post_Weekday + Paid

Coefficients:
```

	Coef.	Std. Error	t	Pr(>|t|)	Lower 95%	Upper 95%
(Intercept)	-15.4458	16.3873	-0.94	0.3464	-47.6447	16.7532
Category	0.38268	1.18089	0.32	0.7460	-1.93761	2.70297
Page_total_likes	0.000230311	0.000178956	1.29	0.1987	-0.000121314	0.000581935
Type: Photo	3.06945	4.86088	0.63	0.5280	-6.48154	12.6204
Type: Status	4.92695	5.79066	0.85	0.3953	-6.45093	16.3048
Type: Video	6.7184	9.3797	0.72	0.4742	-11.7115	25.1483
Post_Month	-1.03585	0.885153	-1.17	0.2425	-2.77506	0.703354
Post_Hour	0.0246612	0.22766	0.11	0.9138	-0.42266	0.471982
Post_Weekday	-0.747369	0.47493	-1.57	0.1162	-1.68054	0.185806
Paid	3.33185	2.15031	1.55	0.1219	-0.893211	7.55692

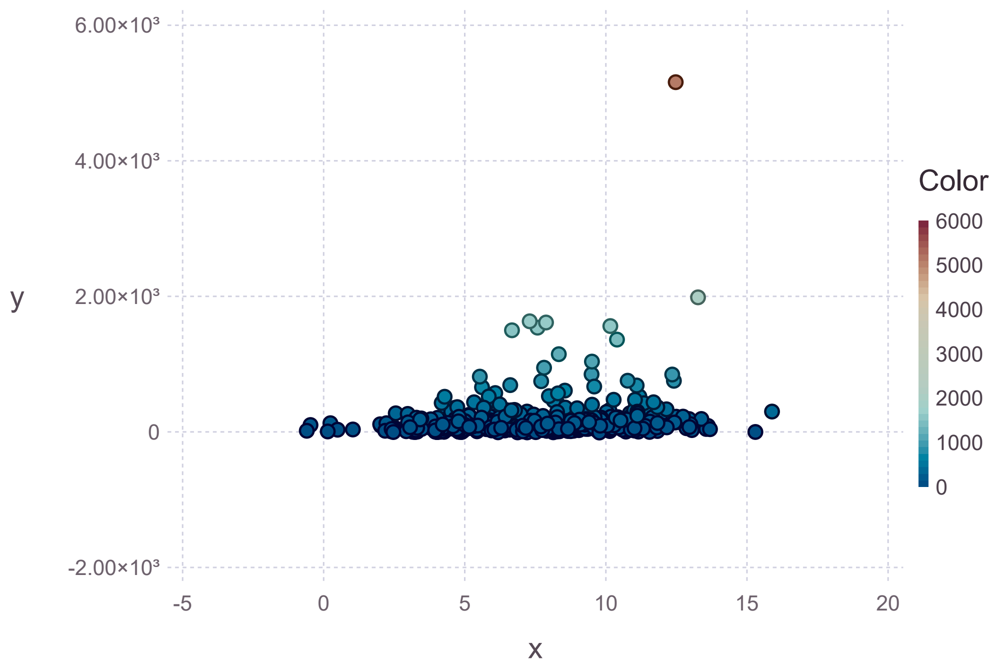
In [267…

```
# predicted values
pred2 = predict(model2)
# residuals
red2 = fb_data[!, :like] - pred2;
```

In [268…

```
# Homoscedasticity assumption
plot(x=pred2, y=red2,color=fb_data[!, :like])
```

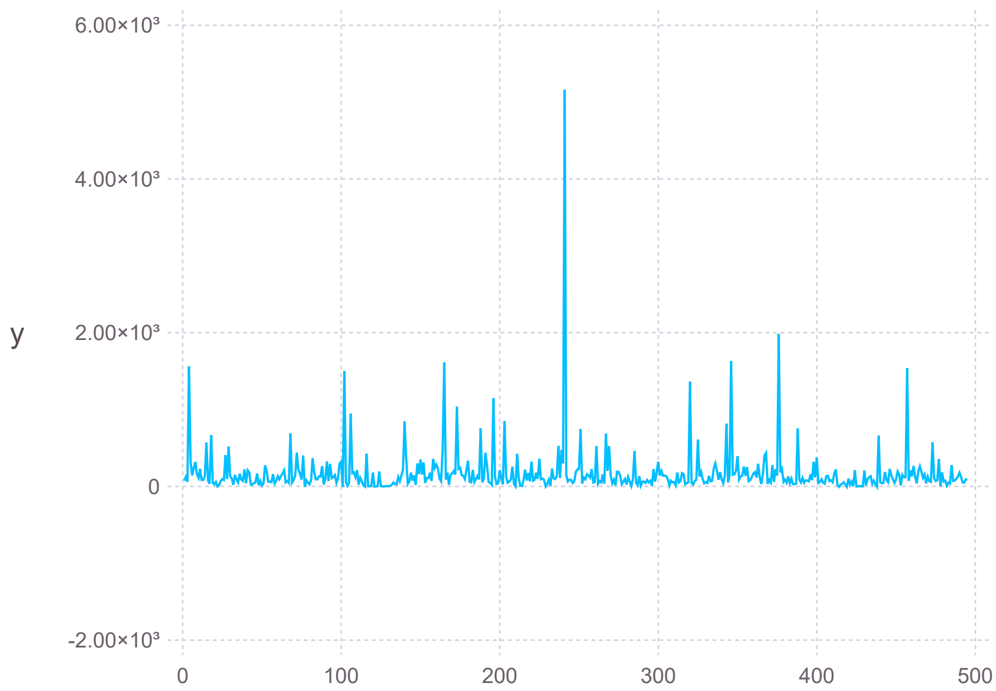
Out[268…



In [269…

```
# independence
plot(y=red2, Geom.line)
```

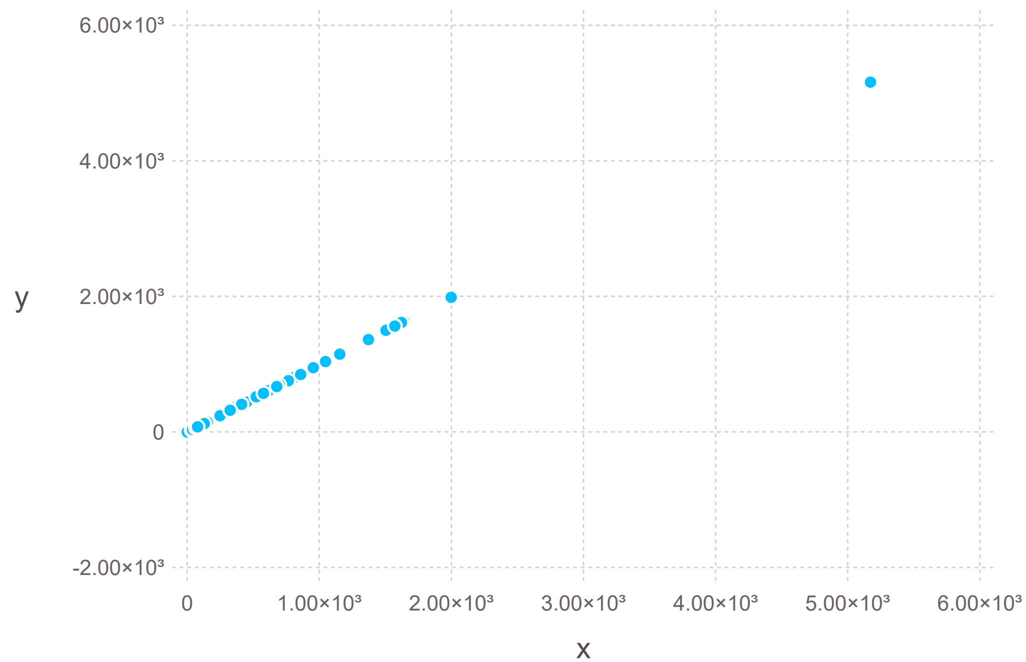
Out[269…



In [270…

```
# Exogeneity assumption
plot(x = fb_data[!, :like], y=red2)
```

Out[270…



In [271…

```
m3 = @formula(share~Category+Page_total_likes+Type+Post_Month+Post_Hour+Post_Weekday+Paid)
```

Out[271…

```
FormulaTerm
Response:
  share(unknown)
Predictors:
  Category(unknown)
  Page_total_likes(unknown)
  Type(unknown)
  Post_Month(unknown)
  Post_Hour(unknown)
  Post_Weekday(unknown)
  Paid(unknown)
```

In [272…

```
model3 = lm(m3, fb_data)
```

Out[272…

```
StatsModels.TableRegressionModel{LinearModel{GLM.LmResp{Vector{Float64}}, GLM.DensePredChol{Float64, CholeskyPivoted{Floa
t64, Matrix{Float64}}}}, Matrix{Float64}}
```

```
share ~ 1 + Category + Page_total_likes + Type + Post_Month + Post_Hour + Post_Weekday + Paid

Coefficients:
──────────────────────────────────────────────────────────────────────────────────────────
                          Coef.     Std. Error      t  Pr(>|t|)        Lower 95%    Upper 95%
──────────────────────────────────────────────────────────────────────────────────────────
(Intercept)        -2.87188      32.4438       -0.09    0.9295  -66.6196       60.8758
Category            6.98957       2.33794       2.99    0.0029    2.39584      11.5833
Page_total_likes    0.000168141   0.000354298   0.47    0.6353   -0.000528007   0.00086429
Type: Photo         8.81314       9.62361       0.92    0.3602  -10.096        27.7223
Type: Status       13.6812       11.4644        1.19    0.2333   -8.84483      36.2072
Type: Video        39.2938       18.57          2.12    0.0349    2.8062       75.7814
Post_Month         -1.13299       1.75243      -0.65    0.5182   -4.57629       2.31031
Post_Hour          -0.471665      0.450722     -1.05    0.2959   -1.35727       0.413945
Post_Weekday       -0.789744      0.94027      -0.84    0.4014   -2.63725       1.05776
Paid                6.65855       4.25719       1.56    0.1185   -1.70626      15.0234
──────────────────────────────────────────────────────────────────────────────────────────
```

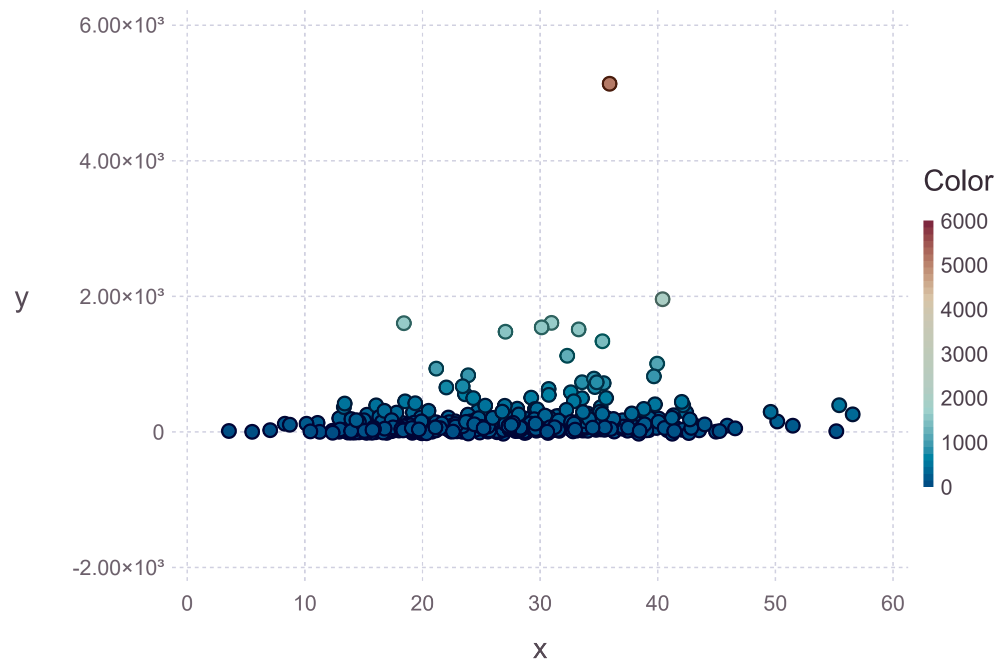
In [273…

```
# predicted values
pred3 = predict(model3)
# residuals
red3 = fb_data[!, :like] - pred3;
```

In [274…

```
# Homoscedasticity assumption
plot(x=pred3, y=red3,color=fb_data[!, :like])
```

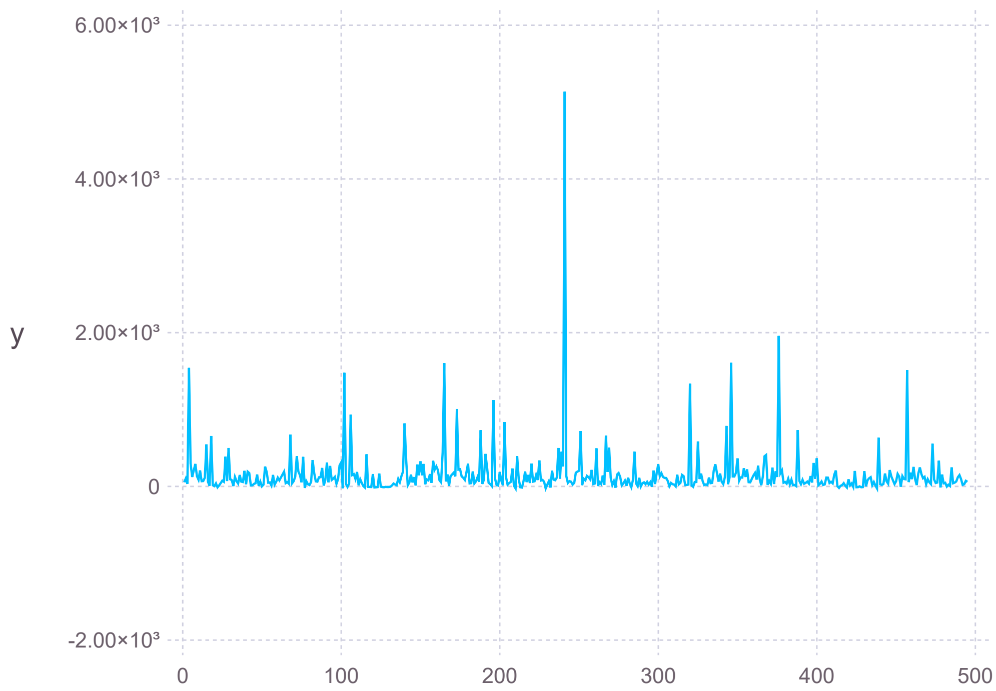
Out[274…



In [275…

```
# independence
plot(y=red3, Geom.line)
```

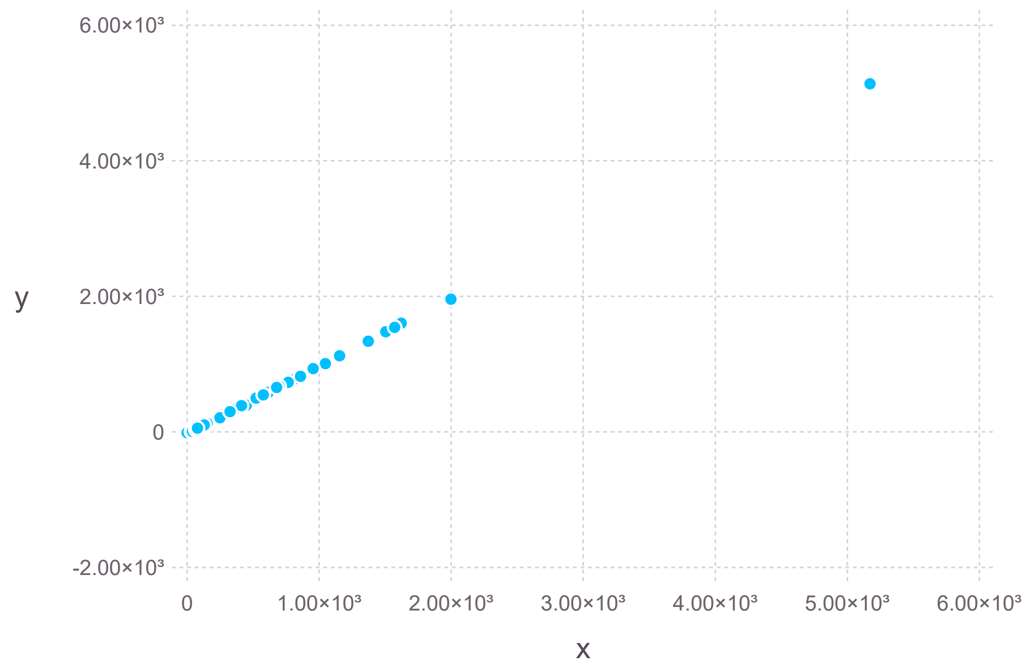
Out[275…



In [276…

```
# Exogeneity assumption
plot(x = fb_data[!, :like], y=red3)
```

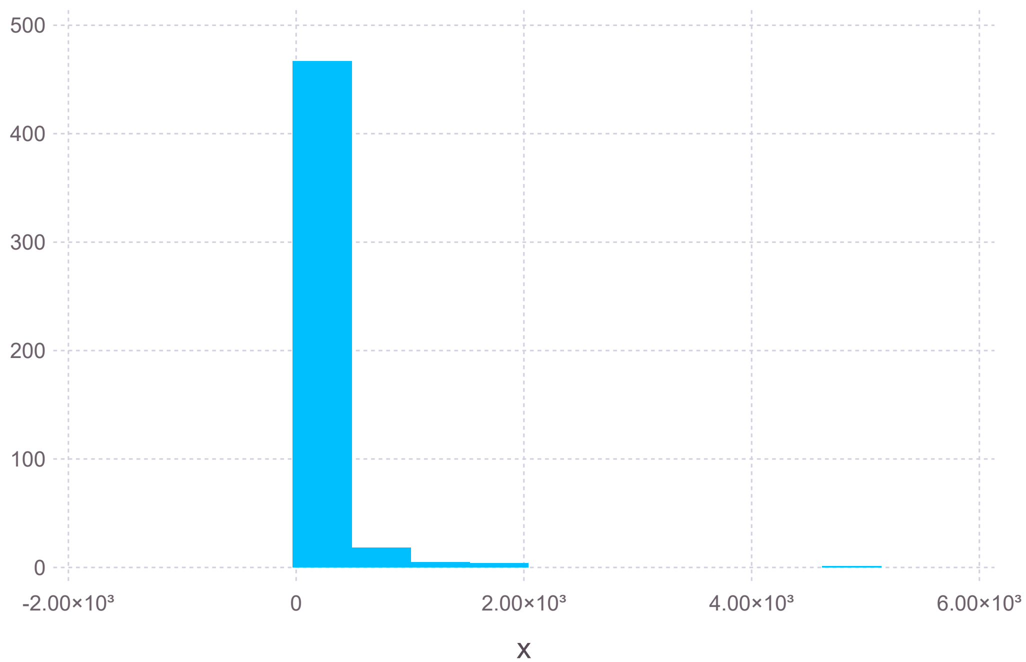
Out[276…



In [277…

```
# error distribution
using Gadfly
plot(x=red3, Geom.histogram(bincount=10))
```

Out[277…



In [278…

```
regtable(model2, model3)
```

```
--------------------------------------
                       comment     share
                      --------  --------
                           (1)       (2)
--------------------------------------
(Intercept)            -15.446    -2.872
                      (16.387)  (32.444)
Category                 0.383   6.990**
                       (1.181)   (2.338)
Page_total_likes         0.000     0.000
                       (0.000)   (0.000)
Type: Photo              3.069     8.813
                       (4.861)   (9.624)
Type: Status             4.927    13.681
                       (5.791)  (11.464)
Type: Video              6.718   39.294*
                       (9.380)  (18.570)
Post_Month              -1.036    -1.133
```

```
                        (0.885)    (1.752)
Post_Hour                 0.025     -0.472
                        (0.228)    (0.451)
Post_Weekday             -0.747     -0.790
                        (0.475)    (0.940)
Paid                      3.332      6.659
                        (2.150)    (4.257)
--------------------------------------
Estimator                   OLS        OLS
--------------------------------------
N                           495        495
R2                        0.018      0.043
--------------------------------------
```

In [279…

```
m2_log = @formula(log1p(comment)~Category+log1p(Page_total_likes)+Type+Post_Month+Post_Hour+Post_Weekday+Paid)
m3_log = @formula(log1p(share)~Category+log1p(Page_total_likes)+Type+Post_Month+Post_Hour+Post_Weekday+Paid)
model2_log = lm(m2_log, fb_data)
```

Out[279…

```
StatsModels.TableRegressionModel{LinearModel{GLM.LmResp{Vector{Float64}}, GLM.DensePredChol{Float64, CholeskyPivoted{Floa
t64, Matrix{Float64}}}}, Matrix{Float64}}

:(log1p(comment)) ~ 1 + Category + :(log1p(Page_total_likes)) + Type + Post_Month + Post_Hour + Post_Weekday + Paid

Coefficients:
──────────────────────────────────────────────────────────────────────────────────────────────
                              Coef.  Std. Error      t  Pr(>|t|)     Lower 95%   Upper 95%
──────────────────────────────────────────────────────────────────────────────────────────────
(Intercept)              -10.2293     10.0858     -1.01    0.3110  -30.0466      9.58791
Category                   0.109983    0.0584973   1.88    0.0607   -0.00495623  0.224923
log1p(Page_total_likes)    0.981862    0.886047    1.11    0.2684   -0.759102    2.72283
Type: Photo                0.288885    0.240698    1.20    0.2306   -0.184054    0.761825
Type: Status               0.651607    0.2868      2.27    0.0235    0.0880819   1.21513
Type: Video                1.18567     0.464549    2.55    0.0110    0.272891    2.09845
Post_Month                -0.035807    0.0386944  -0.93    0.3552   -0.111836    0.0402225
Post_Hour                 -0.00584081  0.0112888  -0.52    0.6051   -0.0280219   0.0163402
Post_Weekday              -0.0385887   0.0235351  -1.64    0.1017   -0.084832    0.00765467
Paid                       0.291965    0.106587    2.74    0.0064    0.0825363   0.501393
──────────────────────────────────────────────────────────────────────────────────────────────
```

In [280…

```
model3_log = lm(m3_log, fb_data)
```

```
StatsModels.TableRegressionModel{LinearModel{GLM.LmResp{Vector{Float64}}, GLM.DensePredChol{Float64, CholeskyPivoted{Floa
```

Out[280…
```
t64, Matrix{Float64}}}}, Matrix{Float64}}

:(log1p(share)) ~ 1 + Category + :(log1p(Page_total_likes)) + Type + Post_Month + Post_Hour + Post_Weekday + Paid

Coefficients:
──────────────────────────────────────────────────────────────────────────────────────────
                              Coef.  Std. Error      t  Pr(>|t|)   Lower 95%    Upper 95%
──────────────────────────────────────────────────────────────────────────────────────────
(Intercept)                7.25892   8.57994       0.85    0.3980  -9.59952    24.1174
Category                   0.431007  0.0497634     8.66    <1e-16   0.333229    0.528786
log1p(Page_total_likes)   -0.448924  0.753756     -0.60    0.5517  -1.92995     1.03211
Type: Photo                0.366645  0.204761      1.79    0.0740  -0.0356827   0.768973
Type: Status               0.64526   0.24398       2.64    0.0084   0.165872    1.12465
Type: Video                1.51791   0.39519       3.84    0.0001   0.741417    2.29441
Post_Month                 0.00506501  0.0329172   0.15    0.8778  -0.0596129   0.0697429
Post_Hour                 -0.029601  0.00960335   -3.08    0.0022  -0.0484703  -0.0107316
Post_Weekday              -0.0342254  0.0200212   -1.71    0.0880  -0.0735644   0.00511364
Paid                       0.0927782  0.0906727    1.02    0.3067  -0.0853816   0.270938
──────────────────────────────────────────────────────────────────────────────────────────
```

In [281…
```
regtable(model2_log, model3_log)
```

```
-----------------------------------------------------
                           log1p(comment)  log1p(share)
                           --------------  ------------
                                      (1)           (2)
-----------------------------------------------------
(Intercept)                       -10.229         7.259
                                 (10.086)       (8.580)
Category                            0.110      0.431***
                                  (0.058)       (0.050)
log1p(Page_total_likes)             0.982        -0.449
                                  (0.886)       (0.754)
Type: Photo                         0.289         0.367
                                  (0.241)       (0.205)
Type: Status                       0.652*       0.645**
                                  (0.287)       (0.244)
Type: Video                        1.186*      1.518***
                                  (0.465)       (0.395)
Post_Month                         -0.036         0.005
                                  (0.039)       (0.033)
Post_Hour                          -0.006      -0.030**
                                  (0.011)       (0.010)
Post_Weekday                       -0.039        -0.034
```

```
                                   (0.024)        (0.020)
Paid                               0.292**          0.093
                                   (0.107)        (0.091)
-------------------------------------------------------
Estimator                              OLS            OLS
-------------------------------------------------------
N                                      495            495
R2                                   0.059          0.204
-------------------------------------------------------
```

In [282…

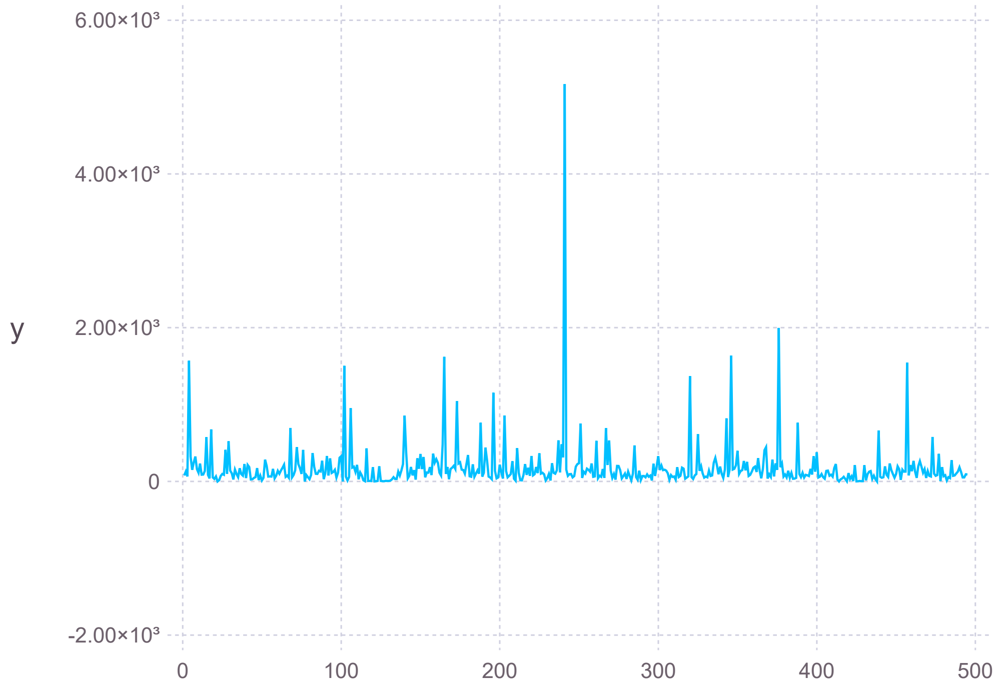
```
# predicted values
pred2_log = predict(model2_log)
# residuals
red2_log = fb_data[!, :like] - pred2_log;

# predicted values
pred3_log = predict(model3_log)
# residuals
red3_log = fb_data[!, :like] - pred3_log;
```

In [283…

```
# independence
plot(y=red2_log, Geom.line)
```

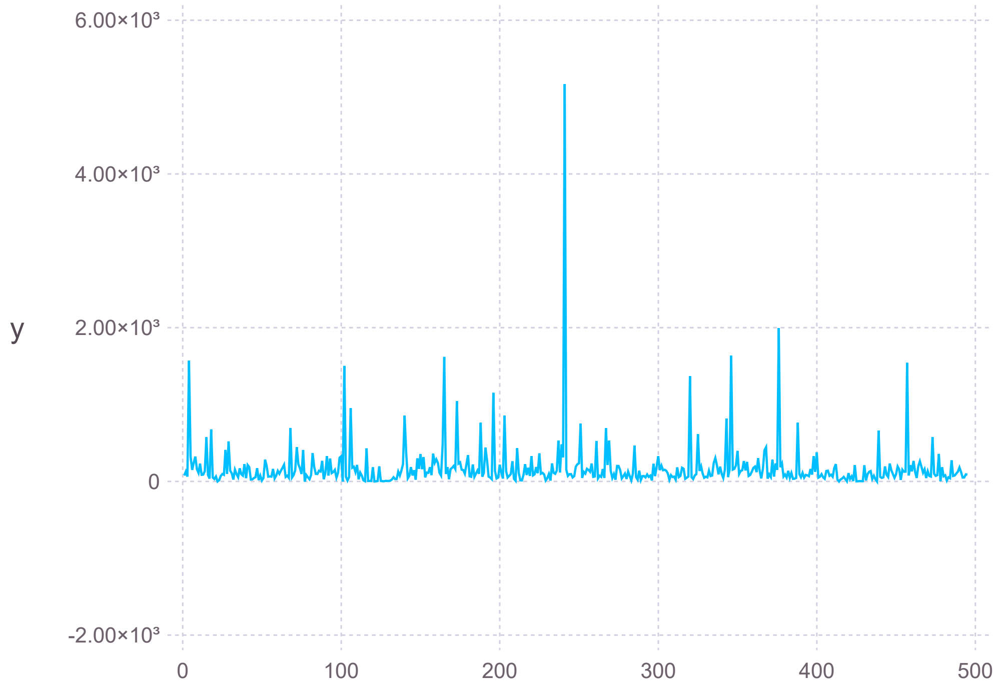
Out[283…



In [284…

```
# independence
plot(y=red3_log, Geom.line)
```

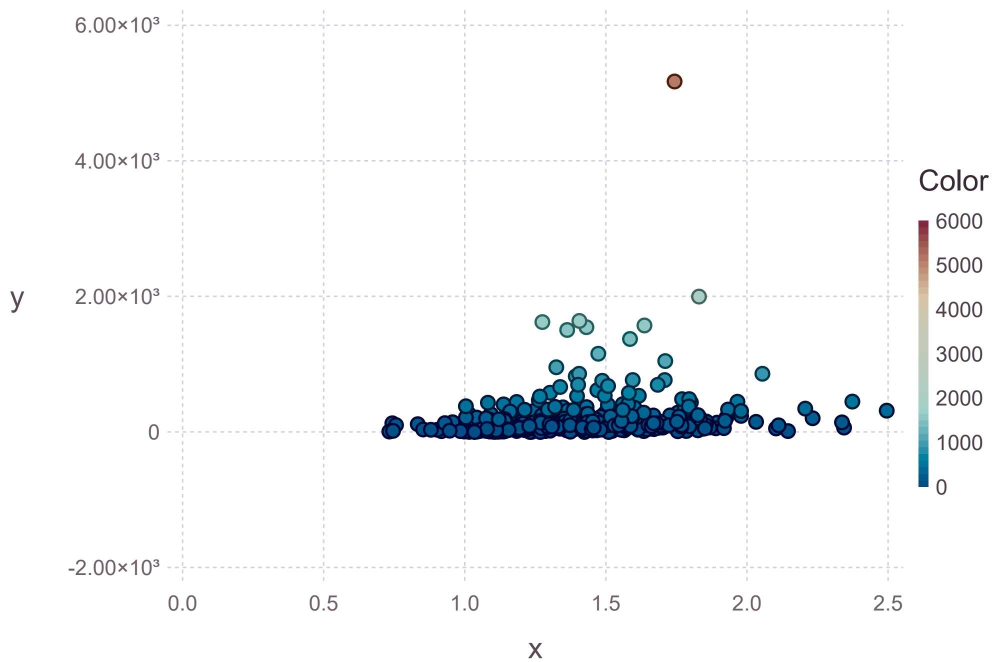
Out[284…



In [285…

```
# Homoscedasticity assumption
plot(x=pred2_log, y=red2_log,color=fb_data[!, :like])
```

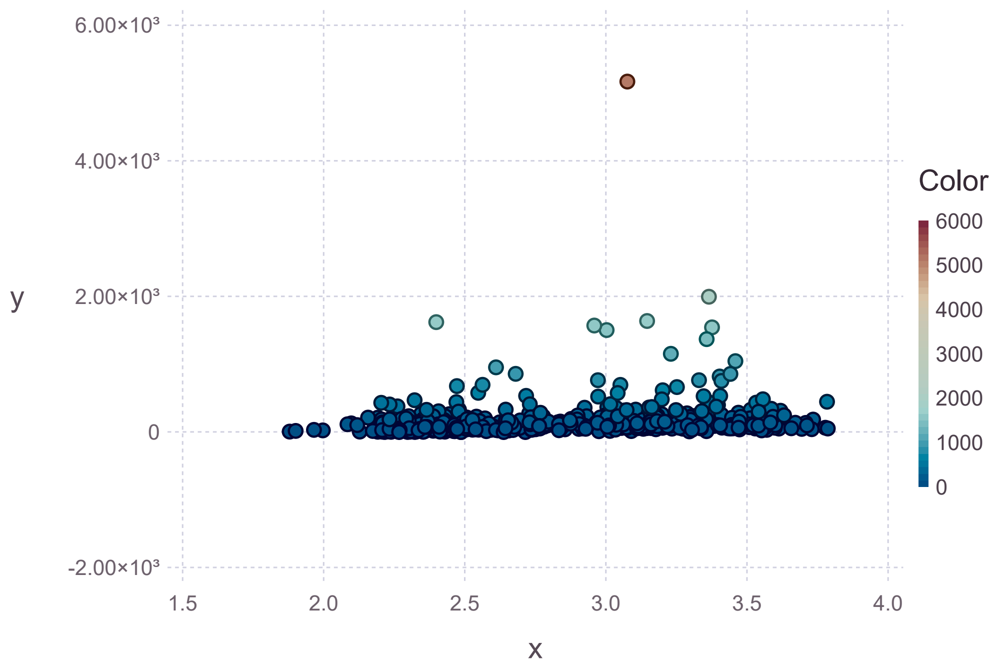
Out[285…



In [286…

```
# Homoscedasticity assumption
plot(x=pred3_log, y=red3_log,color=fb_data[!, :like])
```

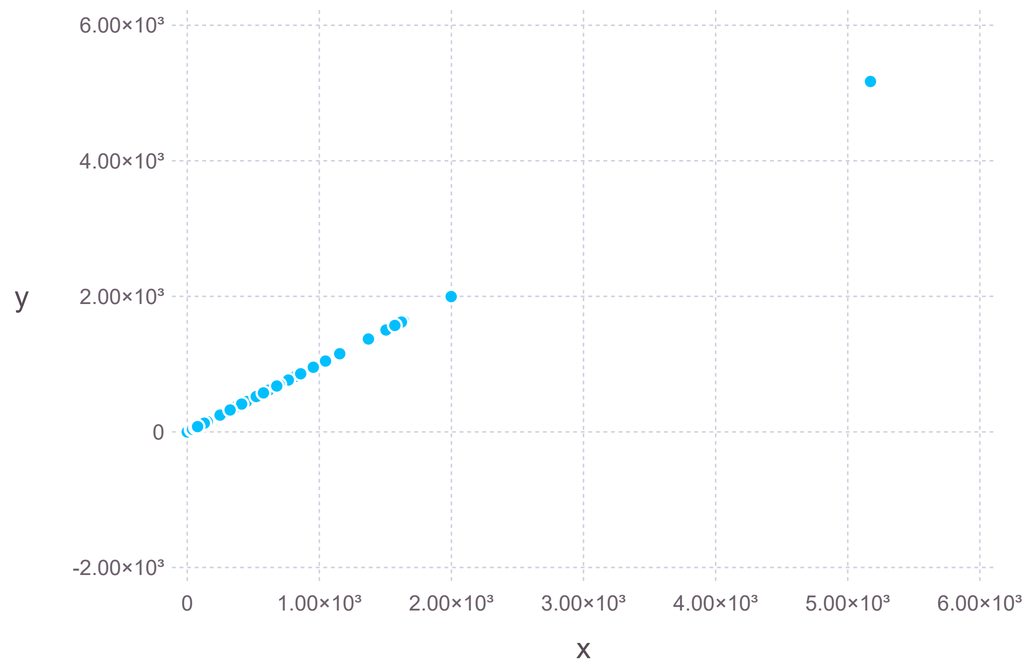
Out[286…



In [287…

```
# Exogeneity assumption
plot(x = fb_data[!, :like], y=red2_log)
```

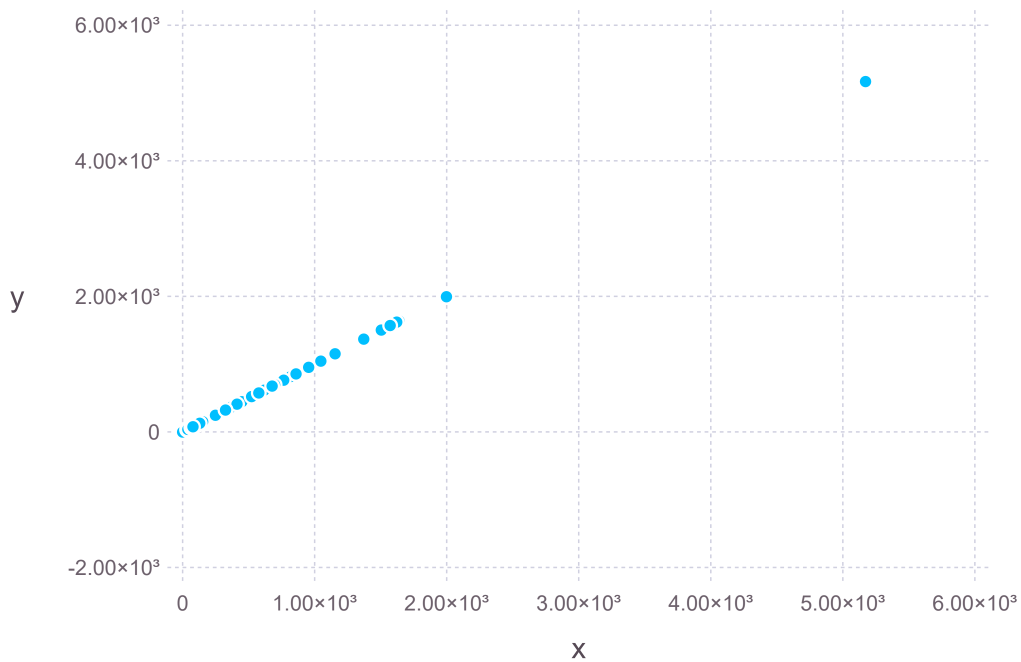
Out[287…



In [288…

```
# Exogeneity assumption
plot(x = fb_data[!, :like], y=red3_log)
```

Out[288…


6.00×10³
4.00×10³
2.00×10³
0
-2.00×10³
y
0
1.00×10³
2.00×10³
3.00×10³
4.00×10³
5.00×10³
6.00×10³
x

In [ ]:

In [ ]: