data overview for summary data

May 5, 2019

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib as plt
%matplotlib inline
import seaborn as sns
```

In [2]:
```
df = pd.read_csv("/Users/MichaelMiao/Documents/career/data.csv")
```

In [3]:
```
df.head(10)
```

Out[3]:

	flight_id	air_temperature	battery_serial_number	body_serial_number
0	16951	20.550000	15SPJJJ09036021	577350132807348254
1	16952	20.500000	15SPJJJ10029029	577209618523054122
2	16954	24.475019	15SPJJJ10012034	577209618523054122
3	16955	27.300000	15SPJJJ10054027	577209618523082792
4	16957	26.950000	15SPJJJ10050049	577348835962150933
5	16959	28.574951	15SPJJJ09018015	577350132840857611
6	16960	27.550000	15SPJJJ09017016	577209618523082792
7	16961	28.250000	15SPJJJ10023027	577350132807348254
8	16962	28.600000	15SPJJJ10052026	577350132840857611
9	16965	32.250000	15SPJJJ10029029	577209618523054122

	commit	launch_airspeed	launch_groundspeed	launch_timestamp
0	5c504d9a16	32.453445	30.164656	2018-09-06 07:43:59 CAT
1	5c504d9a16	32.141209	30.535246	2018-09-06 07:51:49 CAT
2	5c504d9a16	34.701878	29.872613	2018-09-06 09:56:37 CAT
3	5c504d9a16	34.368999	29.877624	2018-09-06 10:27:04 CAT
4	5c504d9a16	32.898979	30.027183	2018-09-06 11:09:39 CAT
5	5c504d9a16	33.258007	30.178811	2018-09-06 11:31:07 CAT
6	5c504d9a16	33.937340	30.063187	2018-09-06 12:55:23 CAT
7	5c504d9a16	33.598980	29.969515	2018-09-06 13:09:51 CAT
8	5c504d9a16	31.639851	30.263741	2018-09-06 13:43:05 CAT
9	5c504d9a16	32.744957	30.354775	2018-09-06 14:56:25 CAT

	preflight_voltage	rel_humidity	static_pressure	wind_direction
0	NaN	74.150000	80662.081736	-49.434555
1	NaN	71.175044	80708.065314	-4.408768
2	NaN	66.374982	80774.270438	-23.458781

```
3                NaN     59.000000     80805.138793      -46.747881
4                NaN     63.900000     80768.969307      -29.293360
5                NaN     65.074951     80621.201026      -68.360838
6                NaN     61.250000     80599.898970      -27.822443
7                NaN     53.500000     80552.491259        7.094333
8                NaN     60.374982     80445.017877      -46.053006
9                NaN     49.600000     80379.646583      -17.594640

   wind_magnitude wing_serial_number
0        1.949338    15SPJJJ11024054
1        0.917357    15SPJJJ09011032
2        3.788383    15SPJJJ09011032
3        3.921605    15SPJJJ11049056
4        2.975881    15SPJJJ09031032
5        2.750346    15SPJJJ11024054
6        1.556340    15SPJJJ09031032
7        2.378607    15SPJJJ11049056
8        1.161924    15SPJJJ09011032
9        2.742027    15SPJJJ11049056
```

In [4]:
```
df.describe() # preflight_voltage has missing vlaues.
```

Out[4]:
```
          flight_id  air_temperature  body_serial_number  launch_airspeed  \
count    447.000000       447.000000        4.470000e+02       447.000000
mean   17373.454139        25.234679        5.773372e+17        31.976493
std      232.833168         4.079718        4.004315e+13         1.759982
min    16951.000000        16.500000        5.772096e+17        28.027149
25%    17170.000000        22.037499        5.773488e+17        30.761058
50%    17359.000000        24.950000        5.773501e+17        31.893215
75%    17590.500000        28.325000        5.773501e+17        33.198513
max    17745.000000        34.600000        5.773501e+17        36.929199

       launch_groundspeed  preflight_voltage  rel_humidity  static_pressure  \
count          447.000000         431.000000    447.000000       447.000000
mean            30.112178          32.145665     56.291622     80456.449107
std              0.371296           0.187159      7.278221       170.113835
min             27.548899          31.544811     35.500000     80010.138720
25%             29.926867          32.055271     51.200017     80323.793028
50%             30.097641          32.189774     56.200000     80445.156986
75%             30.282600          32.268639     61.350000     80590.144407
max             31.205293          32.523598     74.150000     80844.071107

       wind_direction  wind_magnitude
count      447.000000      447.000000
mean       -45.294524        2.359469
std         63.491695        0.996348
min       -176.134898        0.188798
25%        -78.527910        1.703303
```

```
50%        -51.634721        2.307687
75%        -25.951296        3.006968
max        179.700111        7.466193
```

```
In [5]: df['body_serial_number'].value_counts()
```

```
Out[5]: 577350132790558758    69
        577350132807348254    63
        577348835962032159    54
        577350132840857611    45
        577209618523082792    37
        577350132840894487    32
        577350132807389214    26
        577348835878129703    23
        577350132790489124    23
        577348835962155029    22
        577350132807368710    20
        577348835962105883    14
        577348835962150933    13
        577209618523054122     3
        577350132807356452     3
        Name: body_serial_number, dtype: int64
```

```
In [17]: df['battery_serial_number'].value_counts()
```

```
Out[17]: 15SPJJJ10012034    31
         15SPJJJ10029029    27
         15SPJJJ10050016    26
         15SPJJJ09036021    26
         15SPJJJ09018015    24
         15SPJJJ11059037    23
         15SPJJJ10040016    22
         15SPJJJ10022048    21
         15SPJJJ10021047    21
         15SPJJJ09010022    21
         15SPJJJ10019016    20
         15SPJJJ10023027    19
         15SPJJJ10054027    19
         15SPJJJ10060032    18
         15SPJJJ10048030    18
         15SPJJJ10056048    18
         15SPJJJ10007045    15
         15SPJJJ10050049    13
         15SPJJJ10052026    12
         15SPJJJ10030028    11
         15SPJJJ10018016    10
         15SPJJJ10027028    10
         15SPJJJ09017016     8
         15SPJJJ10008029     7
```

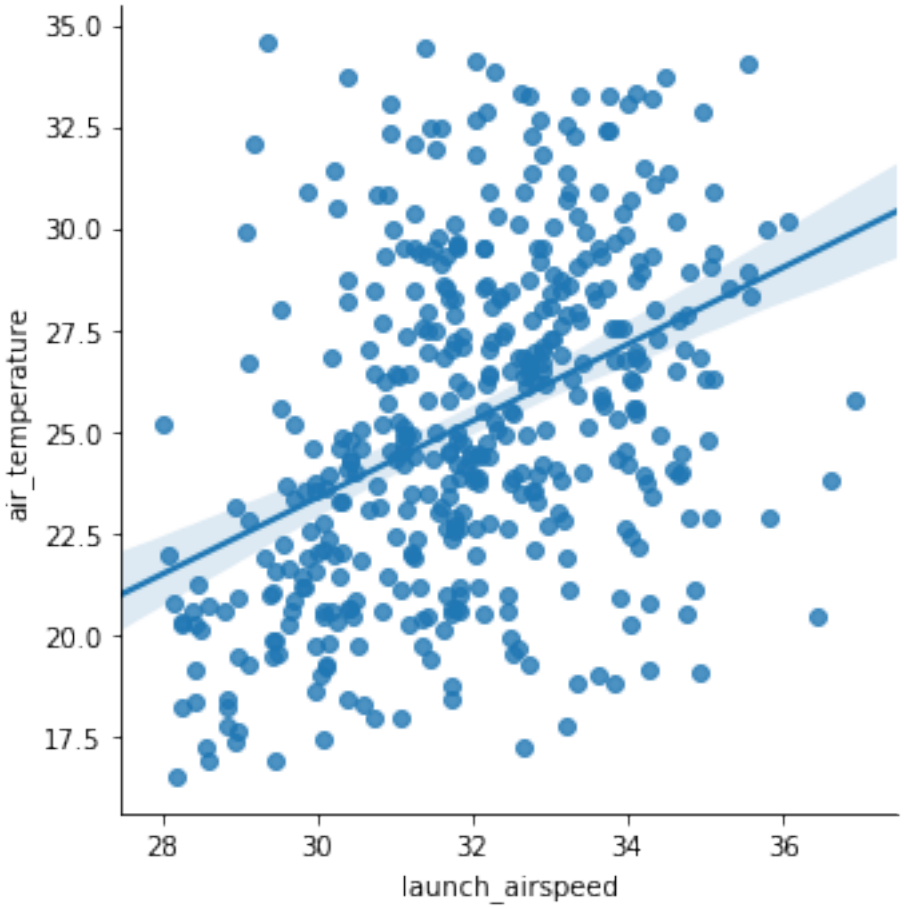
```
          15SPJJJ09013015     4
          15SPJJJ10005031     3
          Name: battery_serial_number, dtype: int64
```

```
In [19]: df['wing_serial_number'].value_counts()
```

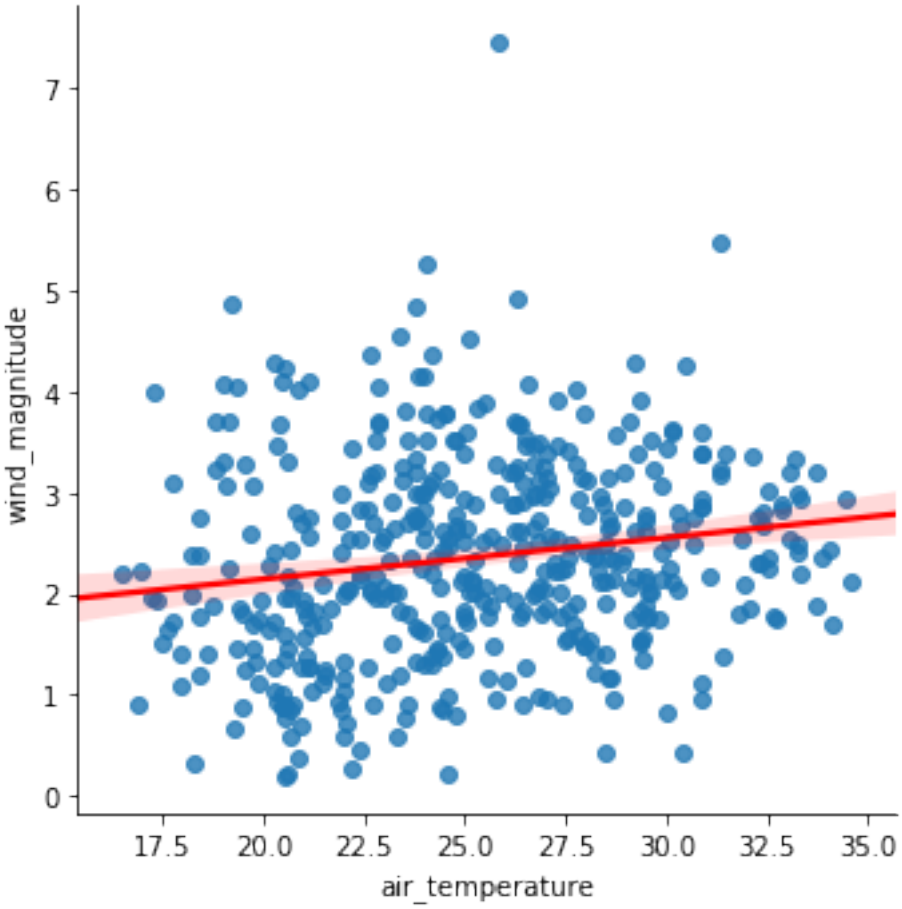
```
Out[19]: 15SPJJJ09008034    65
          15SPJJJ09025064    58
          15SPJJJ09052035    51
          15SPJJJ09024061    45
          15SPJJJ09040032    44
          15SPJJJ09031032    27
          15SPJJJ11049056    24
          15SPJJJ09019061    23
          15SPJJJ09043062    22
          15SPJJJ09036063    17
          15SPJJJ09032034    15
          15SPJJJ09028034    14
          15SPJJJ11024054    13
          15SPJJJ09011032    11
          15SPJJJ09021032     8
          15SPJJJ09010032     5
          15SPJJJ09028064     4
          15SPJJJ11048054     1
          Name: wing_serial_number, dtype: int64
```

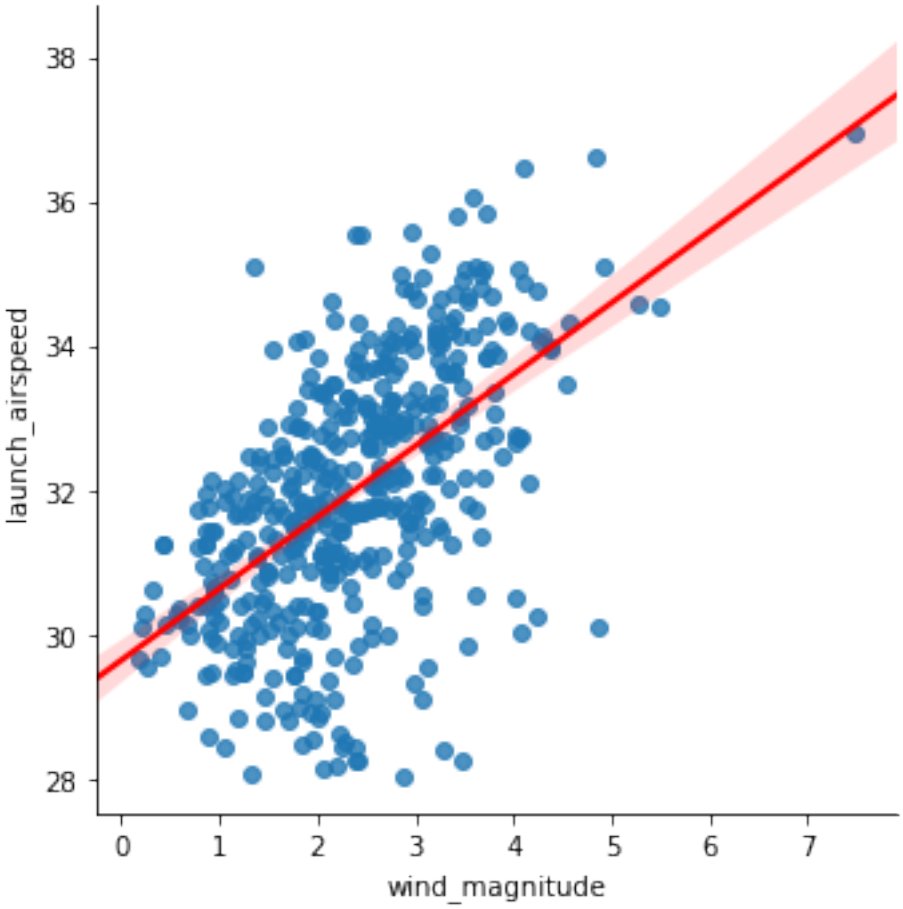
```
In [18]:
```

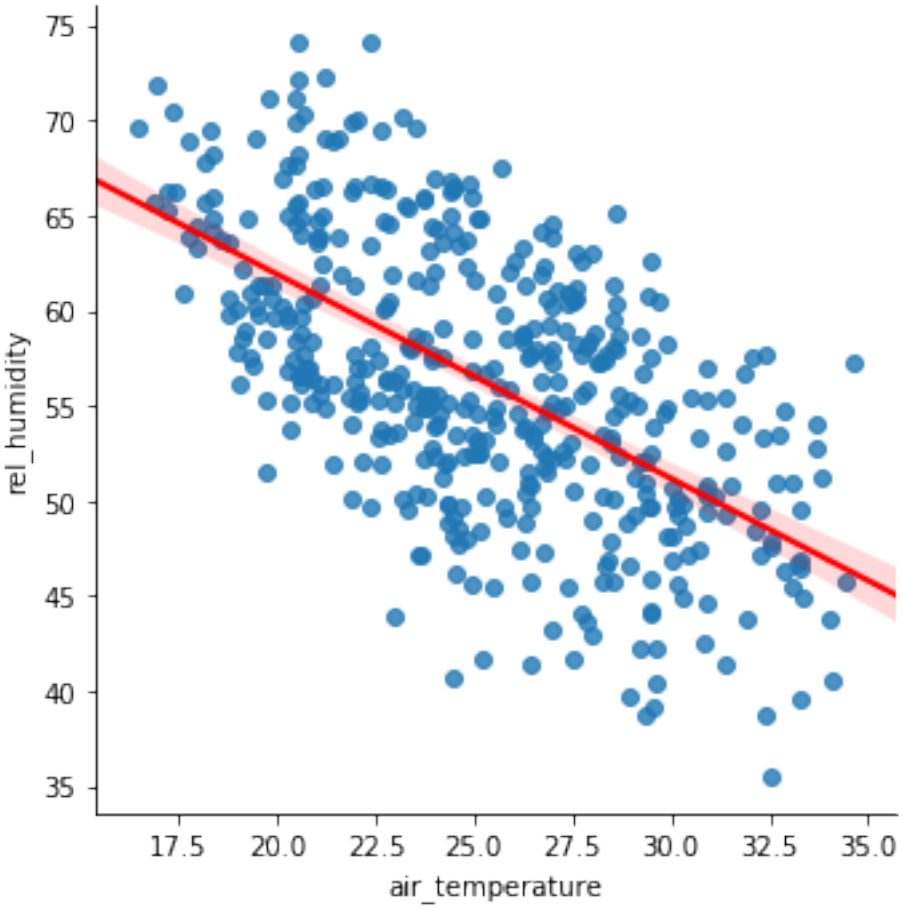
```
Out[18]: <seaborn.axisgrid.FacetGrid at 0x1a129bb0b8>
```

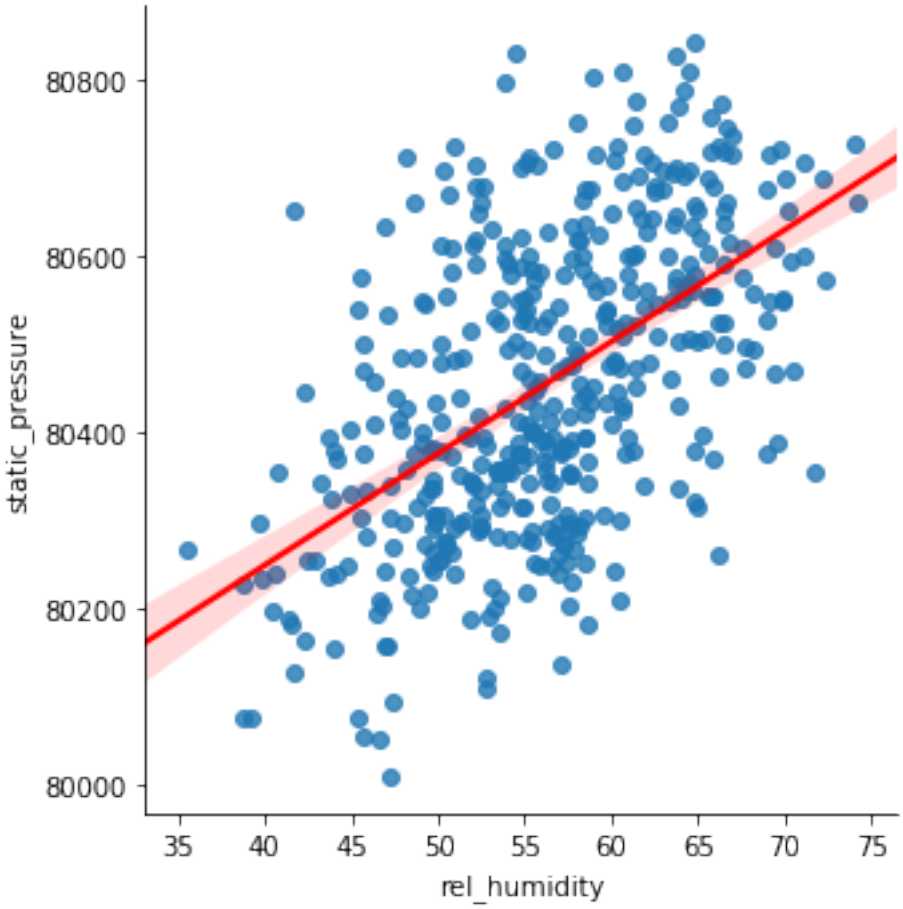


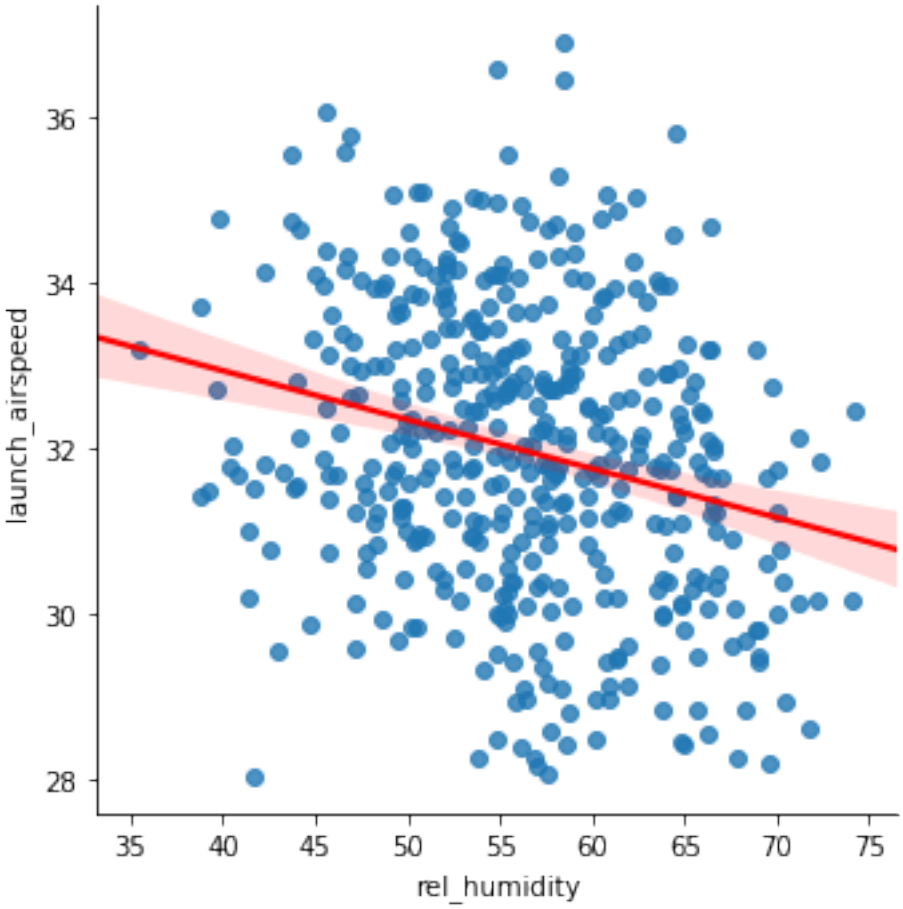

```
In [38]: sns.lmplot(y='wind_magnitude', x= 'air_temperature',data=df,fit_reg=True,line_kws={'co
         sns.lmplot(y='launch_airspeed', x= 'wind_magnitude',data=df,fit_reg=True,line_kws={'co
         sns.lmplot(y='rel_humidity', x='air_temperature',data=df,fit_reg=True,line_kws={'colo
         sns.lmplot(x='rel_humidity', y= 'static_pressure',data=df,fit_reg=True,line_kws={'col
         sns.lmplot(y='launch_airspeed', x= 'rel_humidity',data=df,fit_reg=True,line_kws={'col
         sns.lmplot(y='launch_airspeed', x= 'air_temperature',data=df,fit_reg=True,line_kws={'
         sns.lmplot(y='launch_airspeed', x= 'wind_magnitude',data=df,fit_reg=True,line_kws={'co
         sns.lmplot(y='launch_groundspeed', x= 'wind_magnitude',data=df,fit_reg=True,line_kws=
         sns.lmplot(y='launch_groundspeed', x= 'wind_magnitude',data=df,fit_reg=True,line_kws=
         sns.lmplot(y='static_pressure', x= 'air_temperature',data=df,fit_reg=True,line_kws={'

Out[38]: <seaborn.axisgrid.FacetGrid at 0x1a1343def0>
```

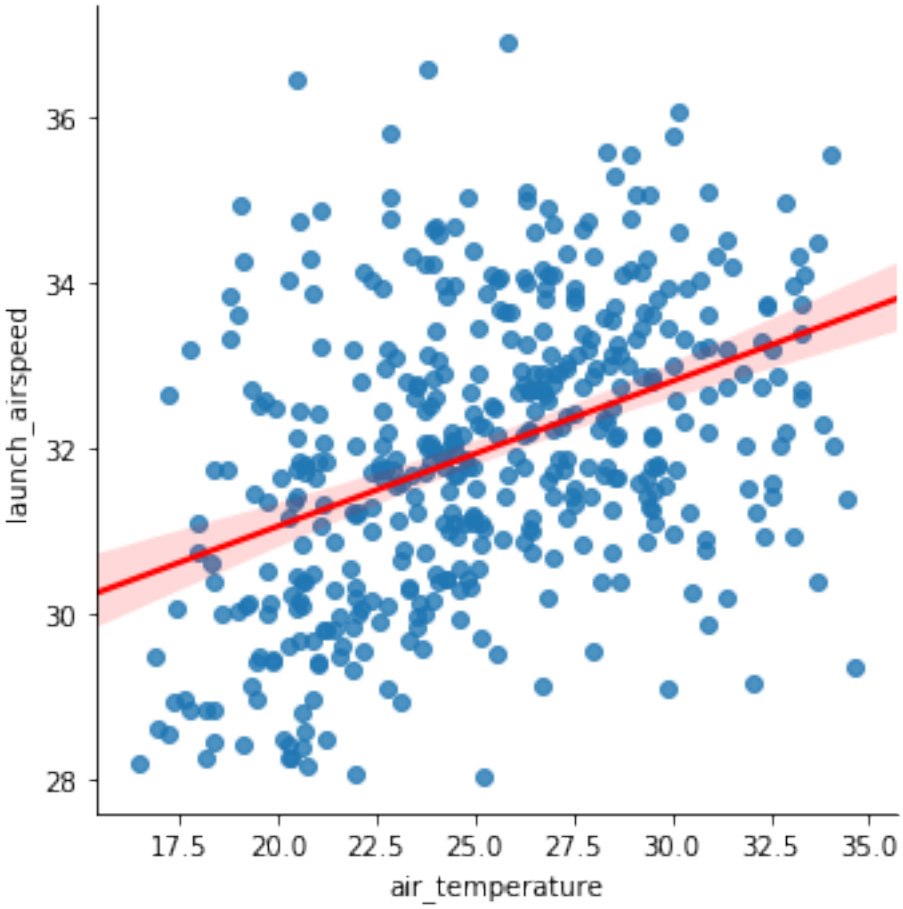

wind_magnitude
0
1
2
3
4
5
6
7
17.5
20.0
22.5
25.0
27.5
30.0
32.5
35.0
air_temperature

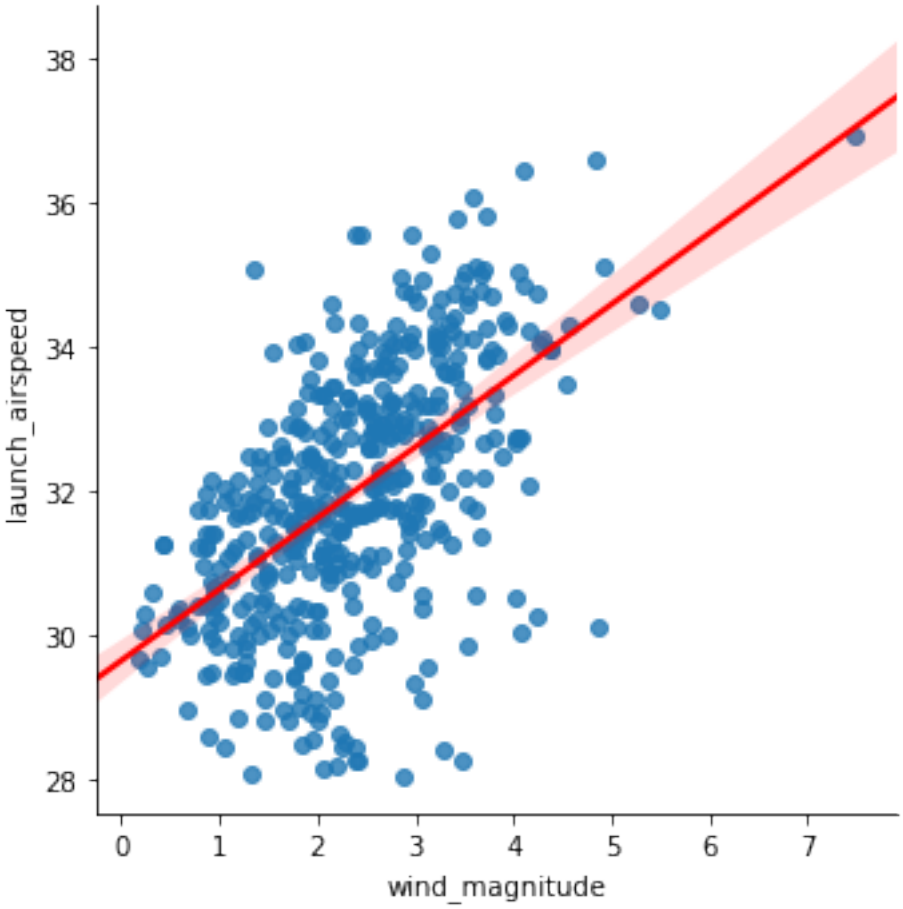

launch_airspeed
wind_magnitude
38
36
34
32
30
28
0
1
2
3
4
5
6
7

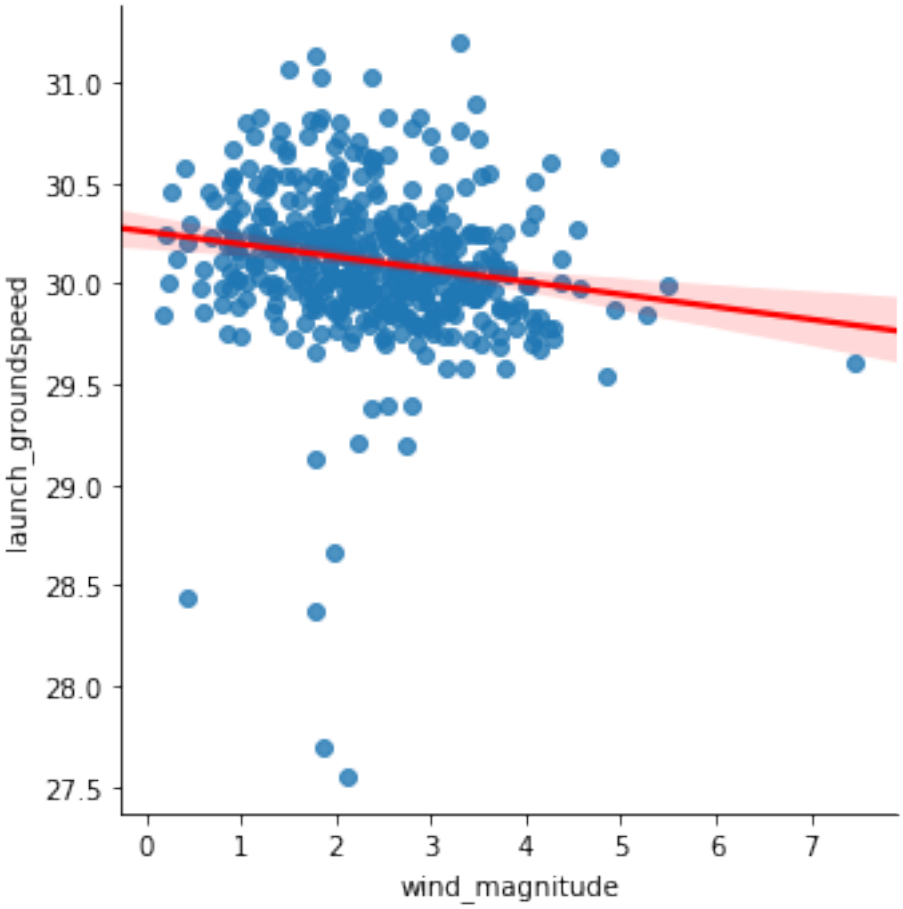

rel_humidity
air_temperature
75
70
65
60
55
50
45
40
35
17.5
20.0
22.5
25.0
27.5
30.0
32.5
35.0

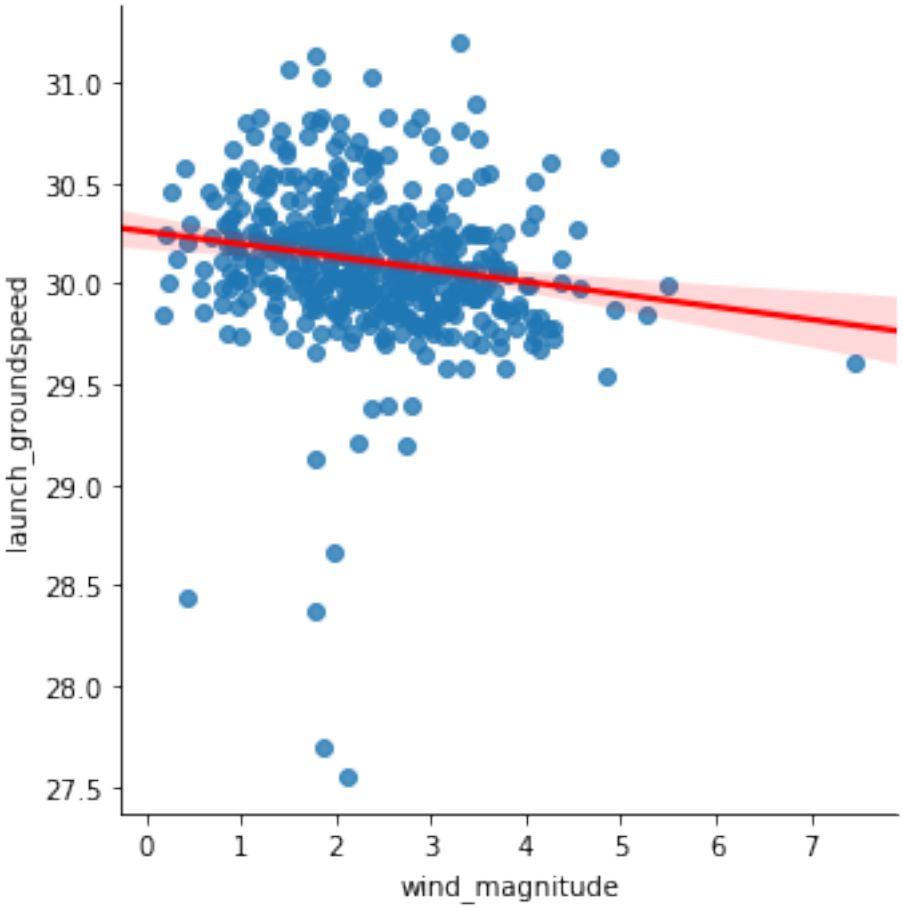

80800
80600
80400
80200
80000
static_pressure
35
40
45
50
55
60
65
70
75
rel_humidity

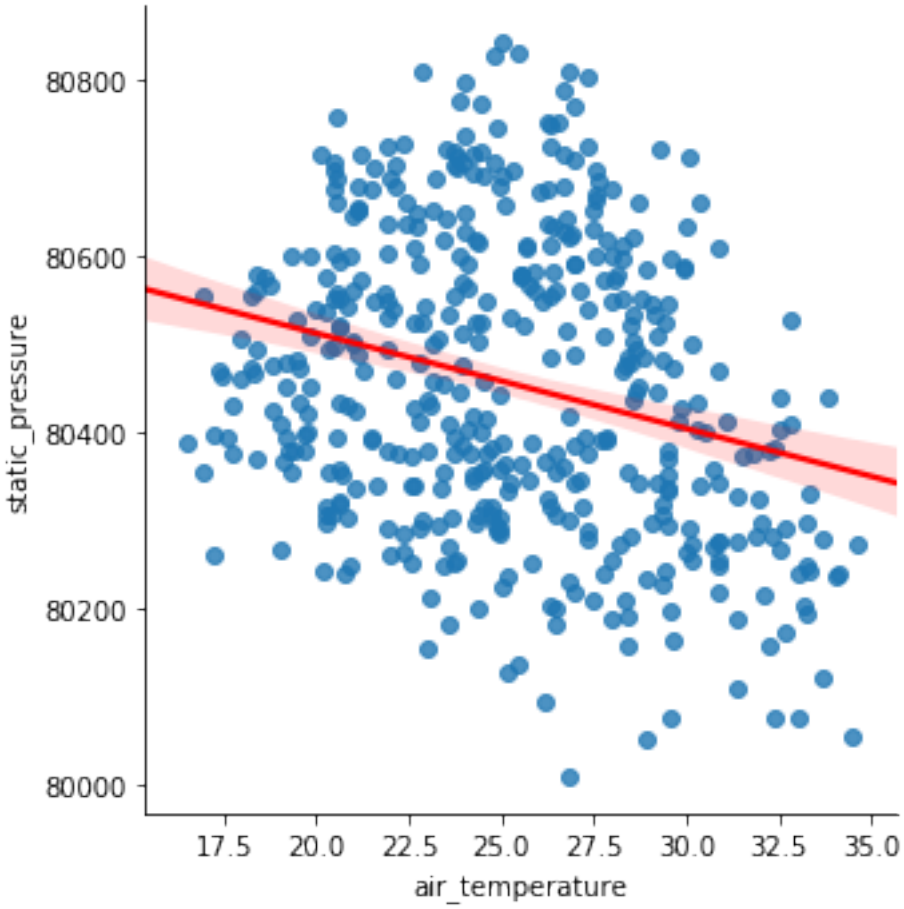

launch_airspeed
rel_humidity
36
34
32
30
28
35
40
45
50
55
60
65
70
75


launch_airspeed
36
34
32
30
28
17.5
20.0
22.5
25.0
27.5
30.0
32.5
35.0
air_temperature


launch_airspeed
wind_magnitude
38
36
34
32
30
28
0
1
2
3
4
5
6
7


launch_groundspeed
31.0
30.5
30.0
29.5
29.0
28.5
28.0
27.5
0
1
2
3
4
5
6
7
wind_magnitude


launch_groundspeed
31.0
30.5
30.0
29.5
29.0
28.5
28.0
27.5
0
1
2
3
4
5
6
7
wind_magnitude



In [43]:
```
# We can explore the linear regression relationship between several variables
#according to the graph.
#The model and more complex exploration
#are written in R markdown file.
```