In [2]:

```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
from sklearn import metrics
import statsmodels.api as sm
from sklearn.preprocessing import LabelEncoder,OneHotEncoder
from sklearn.model_selection import train_test_split
from sklearn.feature_selection import RFE
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
import statsmodels.api as sfa
from scipy import stats
```

In [3]:

```
df1 = pd.read_csv(r"C:\Users\SOURAV CH\Desktop\data science\projects\dataset.csv",';')
pd.set_option('display.max_columns', None) # inorder to axcess all the columns we need to set the max columns to none
pd.set_option('display.max_rows',None)
```

In [4]:

```
df1.head(100)
```

Out[4]:

	Nr	region	subregion	state	vot19_14	turnout14	turnout19	turnout19_14	state_abbrev	debt_2013	debt_20
0	1001	Flensburg, Stadt	1.0	Schleswig-Holstein	-0.003027	0.357400	0.562920	0.205520	KS	16.41	16.
1	1002	Kiel, Landeshauptstadt	1.0	Schleswig-Holstein	0.060316	0.402589	0.588603	0.186015	KS	12.04	12.
2	1003	Lübeck, Hansestadt	1.0	Schleswig-Holstein	0.551817	0.376398	0.546124	0.169726	KS	15.25	15.
3	1004	Neumünster, Stadt	1.0	Schleswig-Holstein	2.392481	0.453649	0.482205	0.028556	KS	16.61	16.
4	1051	Dithmarschen	1.0	Schleswig-Holstein	3.374651	0.397193	0.544108	0.146915	K	12.52	12.
5	1053	Herzogtum Lauenburg	1.0	Schleswig-Holstein	1.646033	0.463842	0.601961	0.138119	K	10.16	10.
6	1054	Nordfriesland	1.0	Schleswig-Holstein	-0.013827	0.411905	0.588692	0.176788	K	10.09	10.
7	1055	Ostholstein	1.0	Schleswig-Holstein	0.252615	0.424788	0.584264	0.159476	K	10.63	10.
8	1056	Pinneberg	1.0	Schleswig-Holstein	0.703336	0.457528	0.629707	0.172180	K	9.67	9.
9	1057	Plön	1.0	Schleswig-Holstein	0.307204	0.469130	0.636865	0.167735	K	8.67	8.
10	1058	Rendsburg-Eckernförde	1.0	Schleswig-Holstein	0.028835	0.459591	0.624910	0.165319	K	9.16	9.
11	1059	Schleswig-Flensburg	1.0	Schleswig-Holstein	-0.508833	0.421130	0.598250	0.177120	K	10.45	10.
12	1060	Segeberg	1.0	Schleswig-Holstein	0.948635	0.423453	0.595325	0.171872	K	10.11	10.
13	1061	Steinburg	1.0	Schleswig-Holstein	1.952564	0.420752	0.577123	0.156371	K	11.43	11.
14	1062	Stormarn	1.0	Schleswig-Holstein	0.523026	0.503786	0.654266	0.150481	K	7.99	7.

	Nr	region	subregion	state	vot19_14	turnout14	turnout19	turnout19_14	state_abbrev	debt_2013	debt_20
15	2000	Hamburg, Freie und Hansestadt	2.0	Hamburg	0.472977	0.435025	0.618714	0.183690	KS	10.92	10.
16	3101	Braunschweig, Stadt	3.0	Niedersachsen	1.221291	0.512913	0.641737	0.128824	KS	10.62	10.
17	3102	Salzgitter, Stadt	3.0	Niedersachsen	9.198355	0.459825	0.511638	0.051813	KS	12.16	12.
18	3103	Wolfsburg, Stadt	3.0	Niedersachsen	3.600941	0.412644	0.564967	0.152323	KS	8.17	7.
19	3151	Gifhorn	3.0	Niedersachsen	4.910447	0.477766	0.607680	0.129914	LK	9.02	8.
20	3153	Goslar	3.0	Niedersachsen	3.475091	0.383769	0.584259	0.200491	LK	12.74	12.
21	3154	Helmstedt	3.0	Niedersachsen	3.786081	0.423608	0.584215	0.160607	LK	10.97	11.
22	3155	Northeim	3.0	Niedersachsen	2.628395	0.481927	0.560951	0.079024	LK	10.80	10.
23	3157	Peine	3.0	Niedersachsen	4.032544	0.468641	0.619507	0.150865	LK	9.78	10.
24	3158	Wolfenbüttel	3.0	Niedersachsen	2.677336	0.525676	0.644388	0.118711	LK	9.13	9.
25	3159	Göttingen	3.0	Niedersachsen	1.762276	0.502767	0.602726	0.099959	LK	9.82	9.
26	3241	Region Hannover	3.0	Niedersachsen	2.288645	0.486508	0.639614	0.153106	LK	12.01	11.
27	3251	Diepholz	3.0	Niedersachsen	2.283782	0.495989	0.621981	0.125992	LK	9.23	9.
28	3252	Hameln-Pyrmont	3.0	Niedersachsen	3.762241	0.491912	0.584282	0.092370	LK	12.23	12.
29	3254	Hildesheim	3.0	Niedersachsen	2.812249	0.484514	0.624283	0.139769	LK	10.98	11.
30	3255	Holzminden	3.0	Niedersachsen	3.587561	0.470233	0.598803	0.128569	LK	11.91	11.
31	3256	Nienburg (Weser)	3.0	Niedersachsen	3.977201	0.431191	0.567161	0.135970	LK	10.26	10.
32	3257	Schaumburg	3.0	Niedersachsen	3.185247	0.506238	0.586828	0.080590	LK	11.54	11.
33	3351	Celle	3.0	Niedersachsen	2.984502	0.465224	0.601866	0.136642	LK	11.90	12.
34	3352	Cuxhaven	3.0	Niedersachsen	3.252164	0.461184	0.618225	0.157041	LK	10.88	10.
35	3353	Harburg	3.0	Niedersachsen	-0.329070	0.522290	0.664795	0.142504	LK	8.12	8.
36	3354	Lüchow-Dannenberg	3.0	Niedersachsen	2.486280	0.529620	0.589581	0.059961	LK	11.55	11.
37	3355	Lüneburg	3.0	Niedersachsen	1.548526	0.531335	0.669420	0.138085	LK	9.97	9.
38	3356	Osterholz	3.0	Niedersachsen	2.277387	0.497100	0.626313	0.129213	LK	8.78	8.
39	3357	Rotenburg (Wümme)	3.0	Niedersachsen	3.029319	0.554893	0.621731	0.066838	LK	9.44	9.
40	3358	Heidekreis	3.0	Niedersachsen	3.528682	0.510705	0.592424	0.081719	LK	10.84	10.
41	3359	Stade	3.0	Niedersachsen	2.274053	0.479329	0.622718	0.143389	LK	9.10	9.
42	3360	Uelzen	3.0	Niedersachsen	2.842799	0.498988	0.617496	0.118508	LK	11.60	11.
43	3361	Verden	3.0	Niedersachsen	2.500429	0.533703	0.642220	0.108517	LK	8.99	8.
44	3401	Delmenhorst, Stadt	3.0	Niedersachsen	5.378949	0.445939	0.526147	0.080208	KS	15.53	15.
45	3402	Emden, Stadt	3.0	Niedersachsen	4.692135	0.359377	0.603340	0.243963	KS	13.57	13.
46	3403	Oldenburg (Oldenburg), Stadt	3.0	Niedersachsen	-0.349140	0.474169	0.648877	0.174708	KS	11.25	10.
47	3404	Osnabrück, Stadt	3.0	Niedersachsen	0.649443	0.509606	0.640557	0.130952	KS	11.28	11.
48	3405	Wilhelmshaven, Stadt	3.0	Niedersachsen	2.308392	0.375954	0.533773	0.157819	KS	16.22	16.
49	3451	Ammerland	3.0	Niedersachsen	0.655735	0.492474	0.648330	0.155856	LK	8.84	8.
50	3452	Aurich	3.0	Niedersachsen	4.669964	0.507975	0.602463	0.094488	LK	10.60	10.
51	3453	Cloppenburg	3.0	Niedersachsen	3.756796	0.506470	0.550193	0.043723	LK	9.11	9.
52	3454	Emsland	3.0	Niedersachsen	2.369881	0.513925	0.628178	0.114253	LK	8.35	8.
53	3455	Friesland	3.0	Niedersachsen	1.762954	0.491948	0.620724	0.128776	LK	10.04	10.
54	3456	Grafschaft Bentheim	3.0	Niedersachsen	1.965304	0.510500	0.671301	0.160801	LK	8.42	8.
55	3457	Leer	3.0	Niedersachsen	3.745500	0.489421	0.557876	0.068455	LK	10.67	10.
56	3458	Oldenburg	3.0	Niedersachsen	1.631730	0.550098	0.629214	0.079116	LK	8.92	8.
57	3459	Osnabrück	3.0	Niedersachsen	2.022159	0.529481	0.626696	0.097214	LK	8.63	8.
58	3460	Vechta	3.0	Niedersachsen	2.555934	0.515049	0.643477	0.128428	LK	7.94	7.
59	3461	Wesermarsch	3.0	Niedersachsen	2.435000	0.423455	0.545466	0.122011	LK	11.10	11.
60	3462	Wittmund	3.0	Niedersachsen	3.795097	0.477998	0.529910	0.051912	LK	10.51	10.

	Nr	region	subregion	state	vot19_14	turnout14	turnout19	turnout19_14	state_abbrev	debt_2013	debt_20
61	4011	Bremen, Stadt	4.0	Bremen	1.393191	0.415112	0.652452	0.237340	KS	12.67	12.6
62	4012	Bremerhaven, Stadt	4.0	Bremen	4.946305	0.345580	0.521454	0.175874	KS	19.84	20.4
63	5111	Düsseldorf, Stadt	5.0	Nordrhein-Westfalen	1.245575	0.538416	0.634823	0.096406	KS	12.47	12.5
64	5112	Duisburg, Stadt	5.0	Nordrhein-Westfalen	4.811571	0.425975	0.500619	0.074644	KS	15.36	15.8
65	5113	Essen, Stadt	5.0	Nordrhein-Westfalen	4.786574	0.473823	0.591842	0.118019	KS	12.80	13.1
66	5114	Krefeld, Stadt	5.0	Nordrhein-Westfalen	1.921189	0.483142	0.579292	0.096150	KS	14.48	14.8
67	5116	Mönchengladbach, Stadt	5.0	Nordrhein-Westfalen	3.548520	0.451435	0.548222	0.096787	KS	15.81	15.8
68	5117	Mülheim an der Ruhr, Stadt	5.0	Nordrhein-Westfalen	2.396875	0.523841	0.626328	0.102487	KS	10.61	10.9
69	5119	Oberhausen, Stadt	5.0	Nordrhein-Westfalen	6.321688	0.455555	0.550390	0.094835	KS	13.53	13.9
70	5120	Remscheid, Stadt	5.0	Nordrhein-Westfalen	3.974041	0.458427	0.563601	0.105175	KS	13.51	13.9
71	5122	Solingen, Klingenstadt	5.0	Nordrhein-Westfalen	3.253057	0.469308	0.573856	0.104548	KS	13.53	14.1
72	5124	Wuppertal, Stadt	5.0	Nordrhein-Westfalen	3.743339	0.481215	0.588017	0.106803	KS	17.89	17.7
73	5154	Kleve	5.0	Nordrhein-Westfalen	1.841317	0.525321	0.602889	0.077568	K	10.21	10.4
74	5158	Mettmann	5.0	Nordrhein-Westfalen	2.260555	0.543117	0.638219	0.095102	K	9.89	10.0
75	5162	Rhein-Kreis Neuss	5.0	Nordrhein-Westfalen	2.299495	0.525470	0.630833	0.105362	K	10.52	10.7
76	5166	Viersen	5.0	Nordrhein-Westfalen	1.844257	0.529101	0.620617	0.091516	K	10.41	10.3
77	5170	Wesel	5.0	Nordrhein-Westfalen	3.810215	0.527408	0.612658	0.085250	K	9.50	9.6
78	5314	Bonn, Stadt	5.0	Nordrhein-Westfalen	0.141390	0.598489	0.694561	0.096072	KS	9.38	9.5
79	5315	Köln, Stadt	5.0	Nordrhein-Westfalen	0.655069	0.531831	0.646348	0.114517	KS	11.80	11.9
80	5316	Leverkusen, Stadt	5.0	Nordrhein-Westfalen	2.770999	0.488064	0.600015	0.111952	KS	11.15	11.3
81	5334	Städteregion Aachen	5.0	Nordrhein-Westfalen	2.891209	0.545318	0.614457	0.069138	SV	11.02	11.0
82	5358	Düren	5.0	Nordrhein-Westfalen	4.444348	0.537839	0.603569	0.065730	K	11.53	11.7
83	5362	Rhein-Erft-Kreis	5.0	Nordrhein-Westfalen	4.198230	0.522865	0.638116	0.115250	K	10.84	10.9
84	5366	Euskirchen	5.0	Nordrhein-Westfalen	3.968619	0.529442	0.612848	0.083406	K	10.86	10.9
85	5370	Heinsberg	5.0	Nordrhein-Westfalen	4.022322	0.542500	0.589030	0.046530	K	11.29	11.4
86	5374	Oberbergischer Kreis	5.0	Nordrhein-Westfalen	3.539610	0.543416	0.610115	0.066699	K	10.00	10.2
87	5378	Rheinisch-Bergischer Kreis	5.0	Nordrhein-Westfalen	1.074745	0.577524	0.680092	0.102567	K	8.77	8.8
88	5382	Rhein-Sieg-Kreis	5.0	Nordrhein-Westfalen	2.271056	0.582665	0.659197	0.076532	K	8.96	8.9
89	5512	Bottrop, Stadt	5.0	Nordrhein-Westfalen	6.703420	0.497681	0.594302	0.096621	KS	11.09	11.4
90	5513	Gelsenkirchen, Stadt	5.0	Nordrhein-Westfalen	8.795457	0.452113	0.512524	0.060411	KS	16.23	16.7
91	5515	Münster, Stadt	5.0	Nordrhein-Westfalen	-0.118364	0.618601	0.736531	0.117930	KS	8.53	8.3
92	5554	Borken	5.0	Nordrhein-Westfalen	2.021129	0.560072	0.650386	0.090314	K	8.79	8.8
93	5558	Coesfeld	5.0	Nordrhein-Westfalen	1.743710	0.601581	0.693412	0.091831	K	7.40	7.4

	Nr	region	subregion	state	vot19_14	turnout14	turnout19	turnout19_14	state_abbrev	debt_2013	debt_20
94	5562	Recklinghausen	5.0	Nordrhein-Westfalen	5.902598	0.493033	0.583251	0.090218	K	12.18	12.
95	5566	Steinfurt	5.0	Nordrhein-Westfalen	2.357753	0.576436	0.649418	0.072981	K	8.85	8.
96	5570	Warendorf	5.0	Nordrhein-Westfalen	2.709944	0.573719	0.649829	0.076110	K	8.74	8.
97	5711	Bielefeld, Stadt	5.0	Nordrhein-Westfalen	2.666558	0.533101	0.637928	0.104827	KS	11.50	11.
98	5754	Gütersloh	5.0	Nordrhein-Westfalen	2.861320	0.539049	0.623341	0.084292	K	8.89	8.
99	5758	Herford	5.0	Nordrhein-Westfalen	4.188082	0.527910	0.592793	0.064883	K	10.62	10.

Simply by looking at the data we can have a first impression of having a high Collinearity as the no of columns are more ,, but we need to further investigate that.

In [7]:

```
print(df1.keys()) # to check the orientation of column names
```

```
Index(['Nr', 'region', 'subregion', 'state', 'vot19_14', 'turnout14',
       'turnout19', 'turnout19_14', 'state_abbrev', 'debt_2013',
       ...
       'f_crime_2015', 'total_suspects_2014', 'foreign_suspects_2014',
       'f_crime_2014', 'total_suspects_2013', 'foreign_suspects_2013',
       'f_crime_2013', 'total_suspects_2012', 'foreign_suspects_2012',
       'f_crime_2012'],
      dtype='object', length=151)
```

In [8]:

```
print(df1.dtypes)
```

```
Nr
int64
region
bject
subregion
float64
state
bject
vot19_14
oat64
turnout14
float64
turnout19
float64
turnout19_14
loat64
state_abbrev
object
debt_2013
float64
debt_2014
float64
debt_2015
float64
debt_2016
float64
debt_2017
float64
debt_2018
float64
ove18_13
oat64
area_2017
float64
population_2017
loat64
```

germans_2017
loat64
foreigners_2017
loat64
population_density_2017
loat64
birth_balance_2017
loat64
net_migration_2017
loat64
age_to_18_2017
loat64
age_18_24_2017
loat64
age_25_34_2017
loat64
age_35_59_2017
loat64
ag4_60_74_2017
loat64
age_75_more_2017
loat64
disposable_inc_2016
int64
gdp_2016
int64
protection_total_2017
loat64
protection_open_2017
loat64
protection_accepted_2017
loat64
protection_rejected_2017
loat64
dwellings_new_2017
loat64
dwellings_2017
loat64
space_per_app_2017
loat64
space_per_inh_2017
loat64
vehicles_2018
loat64
graduates_voc_2017
loat64
Absolventen/Abgänger allgemeinbildender Schulen 2017 - insgesamt ohne Externe (je 1000 Einwohner)
float64
graduates_without_secondary_2017
float64
graduates_lower_secondary_2017
float64
graduates_secondary_2017
loat64
graduates_higher_2017
loat64
child_day_care_2018
loat64
business_reg_2017
loat64
insolvencies_2017
loat64
empl_total_2018
loat64
empl_agr_2018
loat64
empl_manuf_2018
loat64
empl_com_hotel_2018
loat64
empl_service_2018
loat64
empl_oth_service_2018
loat64
hartz_total_2018
loat64
hartz_no_empl_2018

loat64
hartz_foreign_2018
loat64
unempl_total_2019
loat64
unempl_male_2019
loat64
unempl_female_2019
loat64
unempl_15_19_2019
loat64
unempl_55_64_2019
loat64
foreigners_2017_2012
loat64
population_density_2017_2012
loat64
birth_balance_2017_2012
loat64
net_migration_2017_2012
loat64
age_to_18_2017_2012
loat64
age_18_24_2017_2012
loat64
age_25_34_2017_2012
loat64
age_35_59_2017_2012
loat64
ag4_60_74_2017_2012
loat64
age_75_more_2017_2012
loat64
graduates_without_secondary_2017_2012
float64
graduates_lower_secondary_2017_2012
float64
graduates_secondary_2017_2012
loat64
dwellings_new_2017_2012
loat64
dwellings_2017_2012
loat64
vehicles_2018_2013
loat64
business_reg_2017_2012
loat64
insolvencies_2017_2012
loat64
empl_total_2018_2012
loat64
empl_agr_2018_2012
loat64
empl_manuf_2018_2012
loat64
empl_com_hotel_2018_2012
loat64
empl_service_2018_2012
loat64
empl_oth_service_2018_2012
loat64
hartz_total_2018_2013
loat64
hartz_no_empl_2018_2013
loat64
unempl_total_2019_2013
loat64
unempl_male_2019_2013
loat64
unempl_female_2019_2013
loat64
area_2012
float64
population_2012
loat64
male_2012
float64

foreigners_2012
loat64
population_density_2012
loat64
birth_balance_2012
loat64
net_migration_2012
loat64
age_to_18_2012
loat64
age_18_24_2012
loat64
age_25_34_2012
loat64
age_35_59_2012
loat64
ag4_60_74_2012
loat64
age_75_more_2012
loat64
graduates_sec_2012
loat64
graduates_without_secondary_2012
float64
graduates_lower_secondary_2012
float64
graduates_secondary_2012
loat64
graduates_uni_2012
loat64
vehicles_2013
loat64
dwellings_new_2012
loat64
dwellings_2012
loat64
mining_manuf_2012
loat64
trade_tax_per_inh_2012
loat64
business_reg_2012
loat64
business_delist_2012
loat64
insolvencies_2012
loat64
insolvencies_per_1000_2012
loat64
empl_soc_sec_total_2012
loat64
empl_agr_2012
loat64
empl_manuf_2012
loat64
empl_com_hotel_2012
loat64
empl_service_2012
loat64
empl_oth_service_2012
loat64
unempl_total_2013
loat64
unempl_male_2013
loat64
unempl_female_2013
loat64
hartz_total_2013
loat64
hartz_no_empl_2013
loat64
total_suspects_2018
int64
foreign_suspects_2018
int64
f_crime_2018
loat64
total_suspects_2017

```
int64
foreign_suspects_2017
int64
f_crime_2017
loat64
total_suspects_2016
loat64
foreign_suspects_2016
loat64
f_crime_2016
loat64
total_suspects_2015
loat64
foreign_suspects_2015
loat64
f_crime_2015
loat64
total_suspects_2014
loat64
foreign_suspects_2014
loat64
f_crime_2014
loat64
total_suspects_2013
loat64
foreign_suspects_2013
loat64
f_crime_2013
loat64
total_suspects_2012
loat64
foreign_suspects_2012
loat64
f_crime_2012
loat64
dtype: object
```

In [9]:

```
# simply to check the nan values in the columns
df1.isnull().sum().sort_values(ascending = False)
```

Out[9]:

```
empl_agr_2018_2012
6
empl_agr_2018
6
birth_balance_2017_2012
9
empl_manuf_2018
8
empl_manuf_2018_2012
8
birth_balance_2017
7
protection_open_2017
6
empl_service_2018
6
protection_total_2017
6
empl_service_2018_2012
6
protection_accepted_2017
6
protection_rejected_2017
6
insolvencies_per_1000_2012
5
graduates_voc_2017
4
net_migration_2017_2012
3
graduates_uni_2012
2
```

empl_com_hotel_2018_2012
2
graduates_higher_2017
2
empl_com_hotel_2018
2
unempl_total_2013
1
net_migration_2012
1
net_migration_2017
1
unempl_female_2013
1
unempl_male_2013
1
graduates_without_secondary_2017
0
graduates_lower_secondary_2017
0
child_day_care_2018
0
insolvencies_2017
0
graduates_secondary_2017
0
business_reg_2017
0
f_crime_2012
0
empl_total_2018
0
empl_oth_service_2018
0
hartz_total_2018
0
hartz_no_empl_2018
0
hartz_foreign_2018
0
unempl_total_2019
0
unempl_male_2019
0
unempl_female_2019
0
unempl_15_19_2019
0
unempl_55_64_2019
0
foreigners_2017_2012
0
population_density_2017_2012
0
age_to_18_2017_2012
0
Absolventen/Abgänger allgemeinbildender Schulen 2017 - insgesamt ohne Externe (je 1000 Einwohner)
0
dwellings_new_2017
0
vehicles_2018
0
state_abbrev
0
debt_2018
0
debt_2017
0
debt_2016
0
debt_2015
0
debt_2014
0
debt_2013
0
turnout19_14

turnout14
0
0
space_per_inh_2017
0
turnout19
0
turnout14
0
vot19_14
state
subregion
0
region
ove18_13
area_2017
0
population_2017
0
germans_2017
0
foreigners_2017
0
population_density_2017
0
age_to_18_2017
0
age_18_24_2017
0
age_25_34_2017
0
age_35_59_2017
0
ag4_60_74_2017
0
age_75_more_2017
0
disposable_inc_2016
0
gdp_2016
age_25_34_2017_2012
0
dwellings_2017
0
space_per_app_2017
0
age_18_24_2017_2012
0
graduates_secondary_2017_2012
0
age_35_59_2017_2012
0
empl_manuf_2012
0
total_suspects_2018
0
hartz_no_empl_2013
0
hartz_total_2013
0
empl_oth_service_2012
0
empl_service_2012
0
empl_com_hotel_2012
0
empl_agr_2012
0
f_crime_2018
0
empl_soc_sec_total_2012
0
insolvencies_2012
0
business_delist_2012
0
business_reg_2012
0
trade_tax_per_inh_2012

trade_tax_per_inh_2012
0
mining_manuf_2012
0
foreign_suspects_2018
0
total_suspects_2017
0
ag4_60_74_2017_2012
0
total_suspects_2014
0
total_suspects_2012
0
f_crime_2013
0
foreign_suspects_2013
0
total_suspects_2013
0
f_crime_2014
0
foreign_suspects_2014
0
f_crime_2015
0
foreign_suspects_2017
0
foreign_suspects_2015
0
total_suspects_2015
0
f_crime_2016
0
foreign_suspects_2016
0
total_suspects_2016
0
f_crime_2017
0
dwellings_2012
0
dwellings_new_2012
0
vehicles_2013
0
business_reg_2017_2012
0
unempl_total_2019_2013
0
hartz_no_empl_2018_2013
0
hartz_total_2018_2013
0
empl_oth_service_2018_2012
0
empl_total_2018_2012
0
insolvencies_2017_2012
0
vehicles_2018_2013
0
graduates_secondary_2012
0
dwellings_2017_2012
0
dwellings_new_2017_2012
0
foreign_suspects_2012
0
graduates_lower_secondary_2017_2012
0
graduates_without_secondary_2017_2012
0
age_75_more_2017_2012
0
unempl_male_2019_2013
0

```
0
unempl_female_2019_2013
0
area_2012
0
population_2012
0
male_2012
0
foreigners_2012
0
population_density_2012
0
birth_balance_2012
0
age_to_18_2012
0
age_18_24_2012
0
age_25_34_2012
0
age_35_59_2012
0
ag4_60_74_2012
0
age_75_more_2012
0
graduates_sec_2012
0
graduates_without_secondary_2012
0
graduates_lower_secondary_2012
0
Nr
dtype: int64
```

In [39]:

```
df1.describe(include ="all")
# this statement will generally give the overview of data and the information about their behaviou
rs
```

Out[39]:

	Nr	region	subregion	state	vot19_14	turnout14	turnout19	turnout19_14	state_abbrev	debt_2013	debt_2
count	401.000000	401	401.000000	401	401.000000	401.000000	401.000000	401.000000	401	401.000000	401.000
unique	NaN	401	NaN	16	NaN	NaN	NaN	NaN	5	NaN	N
top	NaN	Erfurt, Stadt	NaN	Bayern	NaN	NaN	NaN	NaN	LK	NaN	N
freq	NaN	1	NaN	96	NaN	NaN	NaN	NaN	251	NaN	N
mean	8305.947631	NaN	7.980050	NaN	4.399567	0.471830	0.605913	0.134083	NaN	9.401471	9.502
std	3762.367821	NaN	3.803893	NaN	5.304025	0.071344	0.048212	0.048155	NaN	2.687735	2.729
min	1001.000000	NaN	1.000000	NaN	-2.899875	0.263504	0.476153	0.028556	NaN	3.710000	3.670
25%	5762.000000	NaN	5.000000	NaN	1.091844	0.420203	0.572889	0.095102	NaN	7.440000	7.570
50%	8235.000000	NaN	8.000000	NaN	2.770999	0.475253	0.606205	0.128428	NaN	9.200000	9.220
75%	9676.000000	NaN	9.000000	NaN	4.692135	0.523841	0.639614	0.176785	NaN	10.790000	10.910
max	16077.000000	NaN	16.000000	NaN	21.333849	0.668604	0.743841	0.243963	NaN	19.840000	20.410

In [40]:

```
for i in range(len(df1.index)) :
        print("Nan in rows ",i,":" ,df1.iloc[i].isnull().any().sum())
```

```
Nan in rows  0 : 1
```

```
Nan in rows  1 : 1
Nan in rows  2 : 0
Nan in rows  3 : 0
Nan in rows  4 : 0
Nan in rows  5 : 0
Nan in rows  6 : 0
Nan in rows  7 : 0
Nan in rows  8 : 0
Nan in rows  9 : 0
Nan in rows  10 : 0
Nan in rows  11 : 0
Nan in rows  12 : 0
Nan in rows  13 : 0
Nan in rows  14 : 0
Nan in rows  15 : 0
Nan in rows  16 : 0
Nan in rows  17 : 0
Nan in rows  18 : 1
Nan in rows  19 : 0
Nan in rows  20 : 0
Nan in rows  21 : 0
Nan in rows  22 : 0
Nan in rows  23 : 0
Nan in rows  24 : 0
Nan in rows  25 : 1
Nan in rows  26 : 0
Nan in rows  27 : 0
Nan in rows  28 : 0
Nan in rows  29 : 0
Nan in rows  30 : 0
Nan in rows  31 : 0
Nan in rows  32 : 0
Nan in rows  33 : 0
Nan in rows  34 : 0
Nan in rows  35 : 0
Nan in rows  36 : 0
Nan in rows  37 : 0
Nan in rows  38 : 0
Nan in rows  39 : 0
Nan in rows  40 : 0
Nan in rows  41 : 0
Nan in rows  42 : 0
Nan in rows  43 : 0
Nan in rows  44 : 0
Nan in rows  45 : 0
Nan in rows  46 : 0
Nan in rows  47 : 0
Nan in rows  48 : 0
Nan in rows  49 : 0
Nan in rows  50 : 0
Nan in rows  51 : 0
Nan in rows  52 : 0
Nan in rows  53 : 0
Nan in rows  54 : 0
Nan in rows  55 : 0
Nan in rows  56 : 0
Nan in rows  57 : 0
Nan in rows  58 : 0
Nan in rows  59 : 1
Nan in rows  60 : 1
Nan in rows  61 : 1
Nan in rows  62 : 1
Nan in rows  63 : 0
Nan in rows  64 : 1
Nan in rows  65 : 0
Nan in rows  66 : 0
Nan in rows  67 : 0
Nan in rows  68 : 0
Nan in rows  69 : 1
Nan in rows  70 : 0
Nan in rows  71 : 0
Nan in rows  72 : 0
Nan in rows  73 : 0
Nan in rows  74 : 0
Nan in rows  75 : 0
Nan in rows  76 : 0
Nan in rows  77 : 0
```

```
Nan in rows  77 : 0
Nan in rows  78 : 1
Nan in rows  79 : 0
Nan in rows  80 : 0
Nan in rows  81 : 0
Nan in rows  82 : 0
Nan in rows  83 : 0
Nan in rows  84 : 0
Nan in rows  85 : 0
Nan in rows  86 : 0
Nan in rows  87 : 0
Nan in rows  88 : 0
Nan in rows  89 : 0
Nan in rows  90 : 0
Nan in rows  91 : 0
Nan in rows  92 : 1
Nan in rows  93 : 0
Nan in rows  94 : 0
Nan in rows  95 : 0
Nan in rows  96 : 0
Nan in rows  97 : 0
Nan in rows  98 : 0
Nan in rows  99 : 0
Nan in rows  100 : 0
Nan in rows  101 : 0
Nan in rows  102 : 0
Nan in rows  103 : 1
Nan in rows  104 : 0
Nan in rows  105 : 0
Nan in rows  106 : 0
Nan in rows  107 : 0
Nan in rows  108 : 0
Nan in rows  109 : 0
Nan in rows  110 : 0
Nan in rows  111 : 0
Nan in rows  112 : 0
Nan in rows  113 : 0
Nan in rows  114 : 0
Nan in rows  115 : 0
Nan in rows  116 : 1
Nan in rows  117 : 0
Nan in rows  118 : 0
Nan in rows  119 : 0
Nan in rows  120 : 0
Nan in rows  121 : 0
Nan in rows  122 : 0
Nan in rows  123 : 0
Nan in rows  124 : 0
Nan in rows  125 : 1
Nan in rows  126 : 0
Nan in rows  127 : 0
Nan in rows  128 : 0
Nan in rows  129 : 0
Nan in rows  130 : 0
Nan in rows  131 : 1
Nan in rows  132 : 0
Nan in rows  133 : 0
Nan in rows  134 : 0
Nan in rows  135 : 1
Nan in rows  136 : 0
Nan in rows  137 : 0
Nan in rows  138 : 0
Nan in rows  139 : 0
Nan in rows  140 : 0
Nan in rows  141 : 0
Nan in rows  142 : 0
Nan in rows  143 : 0
Nan in rows  144 : 0
Nan in rows  145 : 0
Nan in rows  146 : 0
Nan in rows  147 : 1
Nan in rows  148 : 0
Nan in rows  149 : 0
Nan in rows  150 : 0
Nan in rows  151 : 1
Nan in rows  152 : 0
Nan in rows  153 : 0
Nan in rows  154 : 0
```

Nan in rows 154 : 0
Nan in rows 155 : 0
Nan in rows 156 : 0
Nan in rows 157 : 0
Nan in rows 158 : 1
Nan in rows 159 : 1
Nan in rows 160 : 0
Nan in rows 161 : 0
Nan in rows 162 : 0
Nan in rows 163 : 0
Nan in rows 164 : 1
Nan in rows 165 : 1
Nan in rows 166 : 0
Nan in rows 167 : 0
Nan in rows 168 : 0
Nan in rows 169 : 0
Nan in rows 170 : 0
Nan in rows 171 : 0
Nan in rows 172 : 1
Nan in rows 173 : 0
Nan in rows 174 : 0
Nan in rows 175 : 0
Nan in rows 176 : 0
Nan in rows 177 : 0
Nan in rows 178 : 0
Nan in rows 179 : 0
Nan in rows 180 : 1
Nan in rows 181 : 0
Nan in rows 182 : 0
Nan in rows 183 : 0
Nan in rows 184 : 0
Nan in rows 185 : 0
Nan in rows 186 : 0
Nan in rows 187 : 0
Nan in rows 188 : 0
Nan in rows 189 : 0
Nan in rows 190 : 0
Nan in rows 191 : 0
Nan in rows 192 : 0
Nan in rows 193 : 0
Nan in rows 194 : 0
Nan in rows 195 : 0
Nan in rows 196 : 0
Nan in rows 197 : 0
Nan in rows 198 : 0
Nan in rows 199 : 0
Nan in rows 200 : 0
Nan in rows 201 : 0
Nan in rows 202 : 0
Nan in rows 203 : 0
Nan in rows 204 : 0
Nan in rows 205 : 1
Nan in rows 206 : 0
Nan in rows 207 : 0
Nan in rows 208 : 0
Nan in rows 209 : 0
Nan in rows 210 : 0
Nan in rows 211 : 0
Nan in rows 212 : 0
Nan in rows 213 : 1
Nan in rows 214 : 0
Nan in rows 215 : 0
Nan in rows 216 : 0
Nan in rows 217 : 1
Nan in rows 218 : 0
Nan in rows 219 : 0
Nan in rows 220 : 0
Nan in rows 221 : 0
Nan in rows 222 : 0
Nan in rows 223 : 0
Nan in rows 224 : 0
Nan in rows 225 : 0
Nan in rows 226 : 0
Nan in rows 227 : 1
Nan in rows 228 : 0
Nan in rows 229 : 0
Nan in rows 230 : 0
Nan in rows 231 : 0

Nan in rows 231 : 0
Nan in rows 232 : 0
Nan in rows 233 : 0
Nan in rows 234 : 0
Nan in rows 235 : 0
Nan in rows 236 : 0
Nan in rows 237 : 0
Nan in rows 238 : 0
Nan in rows 239 : 0
Nan in rows 240 : 0
Nan in rows 241 : 0
Nan in rows 242 : 0
Nan in rows 243 : 0
Nan in rows 244 : 0
Nan in rows 245 : 0
Nan in rows 246 : 0
Nan in rows 247 : 0
Nan in rows 248 : 0
Nan in rows 249 : 0
Nan in rows 250 : 0
Nan in rows 251 : 0
Nan in rows 252 : 0
Nan in rows 253 : 0
Nan in rows 254 : 0
Nan in rows 255 : 0
Nan in rows 256 : 0
Nan in rows 257 : 0
Nan in rows 258 : 0
Nan in rows 259 : 0
Nan in rows 260 : 0
Nan in rows 261 : 0
Nan in rows 262 : 0
Nan in rows 263 : 0
Nan in rows 264 : 0
Nan in rows 265 : 0
Nan in rows 266 : 0
Nan in rows 267 : 0
Nan in rows 268 : 0
Nan in rows 269 : 0
Nan in rows 270 : 0
Nan in rows 271 : 1
Nan in rows 272 : 0
Nan in rows 273 : 1
Nan in rows 274 : 0
Nan in rows 275 : 0
Nan in rows 276 : 0
Nan in rows 277 : 0
Nan in rows 278 : 0
Nan in rows 279 : 0
Nan in rows 280 : 0
Nan in rows 281 : 0
Nan in rows 282 : 0
Nan in rows 283 : 0
Nan in rows 284 : 1
Nan in rows 285 : 0
Nan in rows 286 : 0
Nan in rows 287 : 1
Nan in rows 288 : 0
Nan in rows 289 : 0
Nan in rows 290 : 0
Nan in rows 291 : 0
Nan in rows 292 : 1
Nan in rows 293 : 1
Nan in rows 294 : 0
Nan in rows 295 : 1
Nan in rows 296 : 0
Nan in rows 297 : 0
Nan in rows 298 : 0
Nan in rows 299 : 0
Nan in rows 300 : 0
Nan in rows 301 : 0
Nan in rows 302 : 1
Nan in rows 303 : 0
Nan in rows 304 : 0
Nan in rows 305 : 0
Nan in rows 306 : 0
Nan in rows 307 : 0
Nan in rows 308 : 0

Nan in rows 308 : 0
Nan in rows 309 : 0
Nan in rows 310 : 0
Nan in rows 311 : 0
Nan in rows 312 : 0
Nan in rows 313 : 0
Nan in rows 314 : 0
Nan in rows 315 : 0
Nan in rows 316 : 0
Nan in rows 317 : 0
Nan in rows 318 : 1
Nan in rows 319 : 1
Nan in rows 320 : 1
Nan in rows 321 : 1
Nan in rows 322 : 1
Nan in rows 323 : 1
Nan in rows 324 : 0
Nan in rows 325 : 1
Nan in rows 326 : 1
Nan in rows 327 : 1
Nan in rows 328 : 0
Nan in rows 329 : 0
Nan in rows 330 : 0
Nan in rows 331 : 0
Nan in rows 332 : 0
Nan in rows 333 : 0
Nan in rows 334 : 0
Nan in rows 335 : 1
Nan in rows 336 : 0
Nan in rows 337 : 0
Nan in rows 338 : 0
Nan in rows 339 : 0
Nan in rows 340 : 1
Nan in rows 341 : 0
Nan in rows 342 : 0
Nan in rows 343 : 0
Nan in rows 344 : 0
Nan in rows 345 : 0
Nan in rows 346 : 0
Nan in rows 347 : 0
Nan in rows 348 : 0
Nan in rows 349 : 0
Nan in rows 350 : 0
Nan in rows 351 : 0
Nan in rows 352 : 0
Nan in rows 353 : 0
Nan in rows 354 : 0
Nan in rows 355 : 0
Nan in rows 356 : 0
Nan in rows 357 : 0
Nan in rows 358 : 0
Nan in rows 359 : 0
Nan in rows 360 : 0
Nan in rows 361 : 0
Nan in rows 362 : 0
Nan in rows 363 : 0
Nan in rows 364 : 1
Nan in rows 365 : 0
Nan in rows 366 : 0
Nan in rows 367 : 0
Nan in rows 368 : 0
Nan in rows 369 : 0
Nan in rows 370 : 0
Nan in rows 371 : 0
Nan in rows 372 : 0
Nan in rows 373 : 0
Nan in rows 374 : 0
Nan in rows 375 : 0
Nan in rows 376 : 0
Nan in rows 377 : 0
Nan in rows 378 : 0
Nan in rows 379 : 0
Nan in rows 380 : 1
Nan in rows 381 : 1
Nan in rows 382 : 1
Nan in rows 383 : 0
Nan in rows 384 : 0
Nan in rows 385 : 0

```
Nan in rows  385 : 0
Nan in rows  386 : 0
Nan in rows  387 : 0
Nan in rows  388 : 0
Nan in rows  389 : 0
Nan in rows  390 : 0
Nan in rows  391 : 0
Nan in rows  392 : 0
Nan in rows  393 : 0
Nan in rows  394 : 0
Nan in rows  395 : 1
Nan in rows  396 : 0
Nan in rows  397 : 0
Nan in rows  398 : 0
Nan in rows  399 : 0
Nan in rows  400 : 0
```

In [41]:

```
# accessing the row and column having the null value
df1.iloc[17:19] # then by using the row and column retreving functions we can access the columns h
aving the nan values
```

Out[41]:

	Nr	region	subregion	state	vot19_14	turnout14	turnout19	turnout19_14	state_abbrev	debt_2013	debt_2014	deb
17	3102	Salzgitter, Stadt	3.0	Niedersachsen	9.198355	0.459825	0.511638	0.051813	KS	12.16	12.51	
18	3103	Wolfsburg, Stadt	3.0	Niedersachsen	3.600941	0.412644	0.564967	0.152323	KS	8.17	7.99	

From the above describe function we can see that the mean and median have almost approximately the same values .

- **But as there is a huge difference between then min and max values our data might contain Outliers or some influential points.**
- **There might be some missing values as well.**
- **How do we do that , lets get to know.**

In [42]:

```
# simply to check the nan values in the columns
df1.isnull().sum().sort_values(ascending = False)
```

Out[42]:

```
empl_agr_2018_2012
6
empl_agr_2018
6
birth_balance_2017_2012
9
empl_manuf_2018
8
empl_manuf_2018_2012
8
birth_balance_2017
7
protection_open_2017
6
empl_service_2018
6
protection_total_2017
6
empl_service_2018_2012
6
protection_accepted_2017
6
```

protection_rejected_2017
6
insolvencies_per_1000_2012
5
graduates_voc_2017
4
net_migration_2017_2012
3
graduates_uni_2012
2
empl_com_hotel_2018_2012
2
graduates_higher_2017
2
empl_com_hotel_2018
2
unempl_total_2013
1
net_migration_2012
1
net_migration_2017
1
unempl_female_2013
1
unempl_male_2013
1
graduates_without_secondary_2017
0
graduates_lower_secondary_2017
0
child_day_care_2018
0
insolvencies_2017
0
graduates_secondary_2017
0
business_reg_2017
0
f_crime_2012
0
empl_total_2018
0
empl_oth_service_2018
0
hartz_total_2018
0
hartz_no_empl_2018
0
hartz_foreign_2018
0
unempl_total_2019
0
unempl_male_2019
0
unempl_female_2019
0
unempl_15_19_2019
0
unempl_55_64_2019
0
foreigners_2017_2012
0
population_density_2017_2012
0
age_to_18_2017_2012
0
Absolventen/Abgänger allgemeinbildender Schulen 2017 - insgesamt ohne Externe (je 1000 Einwohner)
0
dwellings_new_2017
0
vehicles_2018
0
state_abbrev
0
debt_2018
0
debt_2017

0
debt_2016
0
debt_2015
0
debt_2014
0
debt_2013
0
turnout19_14
0
space_per_inh_2017
0
turnout19
0
turnout14
0
vot19_14
state
subregion
0
region
ove18_13
area_2017
0
population_2017
0
germans_2017
0
foreigners_2017
0
population_density_2017
0
age_to_18_2017
0
age_18_24_2017
0
age_25_34_2017
0
age_35_59_2017
0
ag4_60_74_2017
0
age_75_more_2017
0
disposable_inc_2016
0
gdp_2016
age_25_34_2017_2012
0
dwellings_2017
0
space_per_app_2017
0
age_18_24_2017_2012
0
graduates_secondary_2017_2012
0
age_35_59_2017_2012
0
empl_manuf_2012
0
total_suspects_2018
0
hartz_no_empl_2013
0
hartz_total_2013
0
empl_oth_service_2012
0
empl_service_2012
0
empl_com_hotel_2012
0
empl_agr_2012
0
f_crime_2018

0
empl_soc_sec_total_2012
0
insolvencies_2012
0
business_delist_2012
0
business_reg_2012
0
trade_tax_per_inh_2012
0
mining_manuf_2012
0
foreign_suspects_2018
0
total_suspects_2017
0
ag4_60_74_2017_2012
0
total_suspects_2014
0
total_suspects_2012
0
f_crime_2013
0
foreign_suspects_2013
0
total_suspects_2013
0
f_crime_2014
0
foreign_suspects_2014
0
f_crime_2015
0
foreign_suspects_2017
0
foreign_suspects_2015
0
total_suspects_2015
0
f_crime_2016
0
foreign_suspects_2016
0
total_suspects_2016
0
f_crime_2017
0
dwellings_2012
0
dwellings_new_2012
0
vehicles_2013
0
business_reg_2017_2012
0
unempl_total_2019_2013
0
hartz_no_empl_2018_2013
0
hartz_total_2018_2013
0
empl_oth_service_2018_2012
0
empl_total_2018_2012
0
insolvencies_2017_2012
0
vehicles_2018_2013
0
graduates_secondary_2012
0
dwellings_2017_2012
0
dwellings_new_2017_2012
0

```
foreign_suspects_2012
0
graduates_lower_secondary_2017_2012
0
graduates_without_secondary_2017_2012
0
age_75_more_2017_2012
0
unempl_male_2019_2013
0
unempl_female_2019_2013
0
area_2012
0
population_2012
0
male_2012
0
foreigners_2012
0
population_density_2012
0
birth_balance_2012
0
age_to_18_2012
0
age_18_24_2012
0
age_25_34_2012
0
age_35_59_2012
0
ag4_60_74_2012
0
age_75_more_2012
0
graduates_sec_2012
0
graduates_without_secondary_2012
0
graduates_lower_secondary_2012
0
Nr
dtype: int64
```

In [43]:

```
# this line will just print the rows having nan values in the colums based on the total nan value
count in it(i.e Ascending order)
df1.assign(Count_NA = lambda x: x.isnull().sum(axis=1)).sort_values('Count_NA', ascending=False)
```

Out[43]:

	Nr	region	subregion	state	vot19_14	turnout14	turnout19	turnout19_14	state_abbrev	debt_2013	de
60	3462	Wittmund	3.0	Niedersachsen	3.795097	0.477998	0.529910	0.051912	LK	10.51	
381	16054	Suhl, Stadt	16.0	Thüringen	10.577550	0.451281	0.564866	0.113586	KS	10.03	
61	4011	Bremen, Stadt	4.0	Bremen	1.393191	0.415112	0.652452	0.237340	KS	12.67	
321	10044	Saarlouis	10.0	Saarland	3.039576	0.542929	0.674194	0.131264	LK	9.81	
380	16053	Jena, Stadt	16.0	Thüringen	5.929222	0.522503	0.649774	0.127272	KS	5.81	
318	10041	Regionalverband Saarbrücken	10.0	Saarland	2.215107	0.482785	0.614821	0.132037	SV	14.13	
165	7318	Speyer, kreisfreie Stadt	7.0	Rheinland-Pfalz	3.542760	0.501333	0.605679	0.104347	KS	10.71	
319	10042	Merzig-Wadern	10.0	Saarland	2.789432	0.603698	0.696526	0.092828	LK	9.80	
320	10043	Neunkirchen	10.0	Saarland	3.421032	0.514172	0.650782	0.136610	LK	11.96	
322	10045	Saarpfalz-Kreis	10.0	Saarland	3.720000	0.573996	0.693711	0.119715	LK	9.08	

322	10045	Saarpfalz-Kreis	10.0	Saarland	3.720000	0.575990	0.695741	0.119743	LK	9.66	
18	3103	Wolfsburg, Stadt	3.0	Niedersachsen	3.600941	0.412644	0.564967	0.152323	KS	8.17	
323	10046	St. Wendel	10.0	Saarland	2.057207	0.654518	0.743841	0.089322	LK	7.87	
325	12051	Brandenburg an der Havel, Stadt	12.0	Brandenburg	9.619840	0.364564	0.486794	0.122230	KS	15.36	
326	12052	Cottbus, Stadt	12.0	Brandenburg	14.885240	0.401710	0.572971	0.171260	KS	11.08	
327	12053	Frankfurt (Oder), Stadt	12.0	Brandenburg	7.942587	0.414892	0.511004	0.096112	KS	12.77	
340	12071	Spree-Neiße	12.0	Brandenburg	21.333849	0.490965	0.613981	0.123016	LK	8.87	
62	4012	Bremerhaven, Stadt	4.0	Bremen	4.946305	0.345580	0.521454	0.175874	KS	19.84	
382	16055	Weimar, Stadt	16.0	Thüringen	8.346655	0.499509	0.630196	0.130687	KS	10.79	
164	7317	Pirmasens, kreisfreie Stadt	7.0	Rheinland-Pfalz	8.605212	0.449477	0.498434	0.048956	KS	17.73	
395	16072	Sonneberg	16.0	Thüringen	19.574445	0.473913	0.574786	0.100873	LK	8.83	
293	9662	Schweinfurt, Stadt	9.0	Bayern	4.226165	0.349141	0.484625	0.135484	KS	9.67	
284	9565	Schwabach, Stadt	9.0	Bayern	0.153551	0.415217	0.597257	0.182040	KS	7.65	
271	9471	Bamberg	9.0	Bayern	4.447954	0.403718	0.604954	0.201236	LK	6.02	
25	3159	Göttingen	3.0	Niedersachsen	1.762276	0.502767	0.602726	0.099959	LK	9.82	
158	7311	Frankenthal (Pfalz), kreisfreie Stadt	7.0	Rheinland-Pfalz	5.367589	0.481309	0.570769	0.089459	KS	12.49	
180	8116	Esslingen	8.0	Baden-Württemberg	2.063479	0.538794	0.661663	0.122870	LK	7.22	
125	6436	Main-Taunus-Kreis	6.0	Hessen	-2.727679	0.496639	0.641114	0.144475	LK	6.88	
335	12066	Oberspreewald-Lausitz	12.0	Brandenburg	18.225485	0.463498	0.567655	0.104157	LK	9.13	
227	9173	Bad Tölz-Wolfratshausen	9.0	Bayern	-2.743907	0.437922	0.633603	0.195680	LK	6.45	
92	5554	Borken	5.0	Nordrhein-Westfalen	2.021129	0.560072	0.650386	0.090314	K	8.79	
213	8415	Reutlingen	8.0	Baden-Württemberg	1.944914	0.501228	0.620379	0.119151	LK	7.24	
135	6611	Kassel, documenta-Stadt	6.0	Hessen	-0.610810	0.399710	0.552927	0.153217	KS	16.32	
205	8316	Emmendingen	8.0	Baden-Württemberg	2.156263	0.538336	0.654813	0.116477	LK	6.98	
302	9678	Schweinfurt	9.0	Bayern	2.185083	0.445546	0.624642	0.179097	LK	4.81	
59	3461	Wesermarsch	3.0	Niedersachsen	2.435000	0.423455	0.545466	0.122011	LK	11.10	
292	9661	Aschaffenburg, Stadt	9.0	Bayern	0.455284	0.342809	0.565078	0.222269	KS	10.64	
273	9473	Coburg	9.0	Bayern	1.580019	0.396983	0.572889	0.175906	LK	7.94	
64	5112	Duisburg, Stadt	5.0	Nordrhein-Westfalen	4.811571	0.425975	0.500619	0.074644	KS	15.36	
159	7312	Kaiserslautern, kreisfreie Stadt	7.0	Rheinland-Pfalz	5.432510	0.436545	0.546304	0.109759	KS	14.60	
103	5774	Paderborn	5.0	Nordrhein-Westfalen	3.032208	0.526459	0.603111	0.076652	K	9.13	
287	9573	Fürth	9.0	Bayern	0.732893	0.440687	0.648255	0.207568	LK	5.78	
116	6411	Darmstadt, Wissenschaftsstadt	6.0	Hessen	-0.339584	0.484667	0.661271	0.176604	KS	9.23	
69	5119	Oberhausen, Stadt	5.0	Nordrhein-Westfalen	6.321688	0.455555	0.550390	0.094835	KS	13.53	
151	7141	Rhein-Lahn-Kreis	7.0	Rheinland-Pfalz	1.895436	0.588342	0.658310	0.069968	LK	10.03	
78	5314	Bonn, Stadt	5.0	Nordrhein-Westfalen	0.141390	0.598489	0.694561	0.096072	KS	9.38	
147	7135	Cochem-Zell	7.0	Rheinland-Pfalz	0.791851	0.635425	0.688101	0.052676	LK	7.69	
131	6532	Lahn-Dill-Kreis	6.0	Hessen	3.324269	0.336814	0.525377	0.188563	LK	9.29	
295	9671	Aschaffenburg	9.0	Bayern	0.707390	0.396348	0.623872	0.227525	LK	6.75	
0	1001	Flensburg, Stadt	1.0	Schleswig-Holstein	-0.003027	0.357400	0.562920	0.205520	KS	16.41	

Nr region subregion state vot19_14 turnout14 turnout19 turnout19_14 state_abbrev debt_2013 de

364	15001	Dessau-Roßlau, Stadt	15.0	Sachsen-Anhalt	12.258134	0.469992	0.545370	0.075378	KS	12.29	
	Nr	**region**	**subregion**	**state**	**vot19_14**	**turnout14**	**turnout19**	**turnout19_14**	**state_abbrev**	**debt_2013**	**de**
217	8425	Alb-Donau-Kreis	8.0	Baden-Württemberg	3.224098	0.575061	0.669995	0.094934	LK	5.72	
172	7335	Kaiserslautern	7.0	Rheinland-Pfalz	6.652091	0.615573	0.673793	0.058220	LK	9.93	
1	1002	Kiel, Landeshauptstadt	1.0	Schleswig-Holstein	0.060316	0.402589	0.588603	0.186015	KS	12.04	
274	9474	Forchheim	9.0	Bayern	1.366896	0.444101	0.638459	0.194357	LK	5.88	
225	9171	Altötting	9.0	Bayern	1.143969	0.369711	0.558422	0.188711	LK	6.62	
279	9479	Wunsiedel i.Fichtelgebirge	9.0	Bayern	3.404176	0.368494	0.545707	0.177212	LK	8.58	
278	9478	Lichtenfels	9.0	Bayern	2.883327	0.398554	0.577273	0.178719	LK	7.01	
277	9477	Kulmbach	9.0	Bayern	1.288777	0.385550	0.574312	0.188762	LK	7.39	
276	9476	Kronach	9.0	Bayern	3.203858	0.373279	0.566682	0.193402	LK	6.86	
275	9475	Hof	9.0	Bayern	2.918919	0.404588	0.569089	0.164502	LK	8.60	
272	9472	Bayreuth	9.0	Bayern	1.091844	0.407352	0.598465	0.191113	LK	6.13	
226	9172	Berchtesgadener Land	9.0	Bayern	0.751892	0.366449	0.577595	0.211146	LK	6.53	
228	9174	Dachau	9.0	Bayern	-0.752110	0.435249	0.651022	0.215773	LK	6.14	
229	9175	Ebersberg	9.0	Bayern	-1.766868	0.476292	0.687105	0.210813	LK	5.57	
230	9176	Eichstätt	9.0	Bayern	2.957736	0.441396	0.636187	0.194791	LK	3.71	
231	9177	Erding	9.0	Bayern	0.025625	0.421546	0.632745	0.211199	LK	6.24	
232	9178	Freising	9.0	Bayern	-1.371404	0.433868	0.636700	0.202832	LK	6.48	
233	9179	Fürstenfeldbruck	9.0	Bayern	-1.805671	0.464263	0.673596	0.209333	LK	6.30	
280	9561	Ansbach, Stadt	9.0	Bayern	2.400414	0.362441	0.539812	0.177371	KS	9.31	
251	9274	Landshut	9.0	Bayern	3.535368	0.398901	0.615971	0.217070	LK	5.52	
235	9181	Landsberg am Lech	9.0	Bayern	-0.857189	0.437467	0.664908	0.227441	LK	5.46	
286	9572	Erlangen-Höchstadt	9.0	Bayern	0.073017	0.470100	0.664929	0.194829	LK	4.63	
215	8417	Zollernalbkreis	8.0	Baden-Württemberg	4.103441	0.488978	0.594303	0.105325	LK	7.55	
216	8421	Ulm, Universitätsstadt	8.0	Baden-Württemberg	0.462649	0.499134	0.640402	0.141267	SK	7.74	
218	8426	Biberach	8.0	Baden-Württemberg	3.384281	0.540911	0.642979	0.102068	LK	6.03	
288	9574	Nürnberger Land	9.0	Bayern	0.951106	0.447498	0.652584	0.205086	LK	6.21	
219	8435	Bodenseekreis	8.0	Baden-Württemberg	1.901173	0.545809	0.667364	0.121555	LK	7.19	
220	8436	Ravensburg	8.0	Baden-Württemberg	1.942558	0.517930	0.640667	0.122737	LK	6.60	
285	9571	Ansbach	9.0	Bayern	1.137629	0.394225	0.591431	0.197206	LK	5.46	
224	9163	Rosenheim, Stadt	9.0	Bayern	-1.414265	0.375733	0.572857	0.197124	KS	9.20	
221	8437	Sigmaringen	8.0	Baden-Württemberg	3.563600	0.537924	0.631844	0.093920	LK	8.88	
222	9161	Ingolstadt, Stadt	9.0	Bayern	2.140122	0.338178	0.529915	0.191737	KS	7.27	
223	9162	München, Landeshauptstadt	9.0	Bayern	-1.808341	0.457512	0.654284	0.196772	KS	8.13	
283	9564	Nürnberg, Stadt	9.0	Bayern	0.360833	0.411337	0.586915	0.175577	KS	10.42	
282	9563	Fürth, Stadt	9.0	Bayern	0.740736	0.364714	0.565319	0.200605	KS	10.79	
281	9562	Erlangen, Stadt	9.0	Bayern	-0.728325	0.482276	0.664848	0.182571	KS	5.99	
234	9180	Garmisch-Partenkirchen	9.0	Bayern	-2.458008	0.427616	0.630824	0.203208	LK	7.23	
270	9464	Hof, Stadt	9.0	Bayern	2.154836	0.330031	0.498835	0.168804	KS	13.39	
252	9275	Passau	9.0	Bayern	2.659498	0.307881	0.528603	0.220723	LK	6.25	
249	9272	Freyung-Grafenau	9.0	Bayern	4.119941	0.264876	0.497436	0.232559	LK	5.58	
260	9371	Amberg-Sulzbach	9.0	Bayern	2.471996	0.406413	0.602980	0.196568	LK	5.52	
246	9262	Passau, Stadt	9.0	Bayern	0.134429	0.376677	0.587239	0.210562	KS	8.82	

	Nr	region	subregion	state	vot19_14	turnout14	turnout19	turnout19_14	state_abbrev	debt_2013	de
259	9363	Weiden i.d.OPf., Stadt	9.0	Bayern	0.149390	0.347440	0.523030	0.175589	KS	10.74	
247	9263	Straubing, Stadt	9.0	Bayern	4.377078	0.318141	0.501187	0.183046	KS	9.99	
290	9576	Roth	9.0	Bayern	1.264727	0.426533	0.630997	0.204465	LK	5.50	
248	9271	Deggendorf	9.0	Bayern	2.858255	0.301582	0.512190	0.210608	LK	6.92	
258	9362	Regensburg, Stadt	9.0	Bayern	-0.481812	0.402922	0.605513	0.202591	KS	9.86	
244	9190	Weilheim-Schongau	9.0	Bayern	-2.899875	0.441989	0.638996	0.197007	LK	6.50	
257	9361	Amberg, Stadt	9.0	Bayern	1.383614	0.346570	0.522411	0.175841	KS	9.15	
250	9273	Kelheim	9.0	Bayern	2.747658	0.375778	0.599959	0.224181	LK	6.66	
256	9279	Dingolfing-Landau	9.0	Bayern	4.841715	0.355939	0.543830	0.187891	LK	6.03	
255	9278	Straubing-Bogen	9.0	Bayern	4.637548	0.396077	0.588264	0.192187	LK	5.09	
254	9277	Rottal-Inn	9.0	Bayern	1.961900	0.330988	0.544201	0.213213	LK	5.81	
253	9276	Regen	9.0	Bayern	4.207446	0.263504	0.476153	0.212649	LK	6.03	
245	9261	Landshut, Stadt	9.0	Bayern	0.608907	0.378682	0.584741	0.206059	KS	8.20	
243	9189	Traunstein	9.0	Bayern	0.787995	0.395355	0.594147	0.198792	LK	5.59	
236	9182	Miesbach	9.0	Bayern	-2.632597	0.420203	0.651096	0.230893	LK	6.89	
240	9186	Pfaffenhofen a.d.Ilm	9.0	Bayern	2.273016	0.437557	0.596782	0.159225	LK	6.22	
237	9183	Mühldorf a.Inn	9.0	Bayern	1.590523	0.372948	0.569719	0.196771	LK	7.73	
238	9184	München	9.0	Bayern	-2.164091	0.493393	0.707205	0.213812	LK	6.09	
239	9185	Neuburg-Schrobenhausen	9.0	Bayern	0.767276	0.422599	0.554723	0.132123	LK	4.95	
269	9463	Coburg, Stadt	9.0	Bayern	-0.344130	0.388115	0.564900	0.176785	KS	9.70	
268	9462	Bayreuth, Stadt	9.0	Bayern	-0.599433	0.379714	0.566443	0.186729	KS	9.41	
267	9461	Bamberg, Stadt	9.0	Bayern	0.787526	0.402236	0.603515	0.201279	KS	8.66	
241	9187	Rosenheim	9.0	Bayern	-0.959835	0.417323	0.630795	0.213472	LK	6.02	
261	9372	Cham	9.0	Bayern	5.156577	0.310838	0.531759	0.220921	LK	5.88	
266	9377	Tirschenreuth	9.0	Bayern	2.524834	0.399240	0.588084	0.188843	LK	5.77	
265	9376	Schwandorf	9.0	Bayern	4.344391	0.359129	0.556446	0.197317	LK	6.79	
242	9188	Starnberg	9.0	Bayern	-2.606281	0.516836	0.719196	0.202360	LK	5.98	
264	9375	Regensburg	9.0	Bayern	2.370208	0.423003	0.625258	0.202255	LK	5.57	
263	9374	Neustadt a.d.Waldnaab	9.0	Bayern	2.148276	0.409582	0.586624	0.177042	LK	5.58	
262	9373	Neumarkt i.d.OPf.	9.0	Bayern	1.694857	0.427384	0.630431	0.203047	LK	5.00	
289	9575	Neustadt a.d.Aisch-Bad Windsheim	9.0	Bayern	1.590525	0.426306	0.612203	0.185897	LK	6.08	
306	9763	Kempten (Allgäu), Stadt	9.0	Bayern	-1.525014	0.358764	0.563103	0.204339	KS	10.33	
291	9577	Weißenburg-Gunzenhausen	9.0	Bayern	0.713554	0.413297	0.585637	0.172340	LK	6.67	
352	14521	Erzgebirgskreis	14.0	Sachsen	16.732576	0.517620	0.643205	0.125585	LK	6.85	
373	15087	Mansfeld-Südharz	15.0	Sachsen-Anhalt	19.206166	0.452224	0.537019	0.084795	LK	11.27	
372	15086	Jerichower Land	15.0	Sachsen-Anhalt	14.359240	0.479496	0.547799	0.068303	LK	11.74	
371	15085	Harz	15.0	Sachsen-Anhalt	12.475906	0.407577	0.535096	0.127519	LK	11.99	
370	15084	Burgenlandkreis	15.0	Sachsen-Anhalt	17.741770	0.458961	0.554543	0.095582	LK	10.74	
369	15083	Börde	15.0	Sachsen-Anhalt	15.039722	0.429236	0.534603	0.105367	LK	10.62	
368	15082	Anhalt-Bitterfeld	15.0	Sachsen-Anhalt	14.517493	0.426044	0.531482	0.105438	LK	11.62	
367	15081	Altmarkkreis Salzwedel	15.0	Sachsen-Anhalt	12.114440	0.453891	0.562001	0.108110	LK	11.85	
366	15003	Magdeburg, Landeshauptstadt	15.0	Sachsen-Anhalt	9.168621	0.385553	0.546574	0.161021	KS	14.64	
365	15002	Halle (Saale), Stadt	15.0	Sachsen-Anhalt	8.993704	0.406083	0.577017	0.170934	KS	17.57	

363	14730	Nordsachsen	14.0	Sachsen	18.555425	0.452866	0.584239	0.131373	LK	10.14	
362	14729	Leipzig	14.0	Sachsen	10.219014	0.480366	0.617409	0.137042	LK	9.32	
361	14713	Leipzig, Stadt	14.0	Sachsen	7.135225	0.426241	0.615974	0.189733	KS	12.91	
360	14628	Sächsische Schweiz-Osterzgebirge	14.0	Sachsen	20.364497	0.540170	0.669254	0.129084	LK	7.44	
359	14627	Meißen	14.0	Sachsen	18.880877	0.506763	0.643710	0.136947	LK	8.02	
358	14626	Görlitz	14.0	Sachsen	20.589026	0.500160	0.634691	0.134530	LK	7.61	
357	14625	Bautzen	14.0	Sachsen	21.251311	0.522053	0.651690	0.129636	LK	7.14	
356	14612	Dresden, Stadt	14.0	Sachsen	10.264851	0.537810	0.686043	0.148233	KS	8.91	
355	14524	Zwickau	14.0	Sachsen	14.512857	0.464305	0.603786	0.139481	LK	8.62	
354	14523	Vogtlandkreis	14.0	Sachsen	13.706254	0.482843	0.611456	0.128613	LK	9.35	
374	15088	Saalekreis	15.0	Sachsen-Anhalt	17.588113	0.447176	0.571108	0.123932	LK	11.30	
375	15089	Salzlandkreis	15.0	Sachsen-Anhalt	16.346450	0.405973	0.500385	0.094412	LK	12.33	
376	15090	Stendal	15.0	Sachsen-Anhalt	14.090513	0.427994	0.543683	0.115690	LK	11.92	
390	16067	Gotha	16.0	Thüringen	15.260989	0.519206	0.598638	0.079432	LK	9.66	
399	16076	Greiz	16.0	Thüringen	15.881514	0.553222	0.641391	0.088169	LK	7.42	
398	16075	Saale-Orla-Kreis	16.0	Thüringen	18.719615	0.541176	0.634263	0.093087	LK	7.48	
397	16074	Saale-Holzland-Kreis	16.0	Thüringen	13.389590	0.584836	0.660550	0.075714	LK	7.40	
396	16073	Saalfeld-Rudolstadt	16.0	Thüringen	18.712021	0.508175	0.609925	0.101750	LK	7.45	
394	16071	Weimarer Land	16.0	Thüringen	16.939578	0.565730	0.657761	0.092031	LK	9.30	
393	16070	Ilm-Kreis	16.0	Thüringen	15.364231	0.528120	0.634263	0.106144	LK	9.18	
392	16069	Hildburghausen	16.0	Thüringen	17.625163	0.524583	0.646276	0.121694	LK	8.01	
391	16068	Sömmerda	16.0	Thüringen	17.965880	0.537037	0.611854	0.074817	LK	9.03	
389	16066	Schmalkalden-Meiningen	16.0	Thüringen	15.582975	0.521090	0.611689	0.090600	LK	8.92	
377	15091	Wittenberg	15.0	Sachsen-Anhalt	15.177903	0.467126	0.575578	0.108452	LK	9.96	
388	16065	Kyffhäuserkreis	16.0	Thüringen	18.334998	0.516722	0.594564	0.077842	LK	10.88	
387	16064	Unstrut-Hainich-Kreis	16.0	Thüringen	16.973704	0.496123	0.594907	0.098784	LK	9.80	
386	16063	Wartburgkreis	16.0	Thüringen	16.664091	0.544328	0.625806	0.081478	LK	7.53	
385	16062	Nordhausen	16.0	Thüringen	16.782858	0.485793	0.588523	0.102730	LK	9.47	
384	16061	Eichsfeld	16.0	Thüringen	12.764963	0.573239	0.658565	0.085326	LK	6.19	
383	16056	Eisenach, Stadt	16.0	Thüringen	14.983195	0.516287	0.595492	0.079205	KS	11.87	
379	16052	Gera, Stadt	16.0	Thüringen	18.453636	0.435021	0.567636	0.132615	KS	11.62	
378	16051	Erfurt, Stadt	16.0	Thüringen	9.566501	0.479950	0.607435	0.127485	KS	11.48	
353	14522	Mittelsachsen	14.0	Sachsen	19.040197	0.516281	0.642359	0.126078	LK	8.14	
351	14511	Chemnitz, Stadt	14.0	Sachsen	13.811831	0.444628	0.624049	0.179421	KS	10.22	
294	9663	Würzburg, Stadt	9.0	Bayern	-1.054141	0.417390	0.631703	0.214313	KS	7.01	
350	13076	Ludwigslust-Parchim	13.0	Mecklenburg-Vorpommern	11.293800	0.512948	0.604610	0.091663	LK	9.59	
316	9779	Donau-Ries	9.0	Bayern	0.388621	0.405818	0.590375	0.184557	LK	5.17	
315	9778	Unterallgäu	9.0	Bayern	0.395259	0.388968	0.588668	0.199699	LK	6.03	
314	9777	Ostallgäu	9.0	Bayern	-0.724483	0.408860	0.615390	0.206531	LK	6.31	
313	9776	Lindau (Bodensee)	9.0	Bayern	-0.219386	0.399144	0.617607	0.218463	LK	7.63	
312	9775	Neu-Ulm	9.0	Bayern	-0.702398	0.372377	0.559146	0.186769	LK	6.94	
311	9774	Günzburg	9.0	Bayern	0.486533	0.351691	0.542050	0.190359	LK	6.66	
310	9773	Dillingen a.d.Donau	9.0	Bayern	3.420186	0.366263	0.547142	0.180878	LK	6.48	
309	9772	Augsburg	9.0	Bayern	-1.207842	0.400863	0.602695	0.201832	LK	6.16	
308	9771	Aichach-Friedberg	9.0	Bayern	-0.425402	0.412541	0.631524	0.218983	LK	5.27	
307	9764	Memmingen, Stadt	9.0	Bayern	-1.073392	0.373119	0.573507	0.200388	KS	10.49	

305	9762	Kaufbeuren, Stadt	9.0	Bayern	0.433779	0.351161	0.523863	0.172701	KS	10.66	
304	9761	Augsburg, Stadt	9.0	Bayern	-1.444398	0.363072	0.555695	0.192622	KS	11.04	
	Nr	**region**	**subregion**	**state**	**vot19_14**	**turnout14**	**turnout19**	**turnout19_14**	**state_abbrev**	**debt_2013**	**de**
303	9679	Würzburg	9.0	Bayern	0.080474	0.475480	0.680799	0.205319	LK	5.15	
301	9677	Main-Spessart	9.0	Bayern	-0.057201	0.421136	0.631879	0.210743	LK	5.37	
300	9676	Miltenberg	9.0	Bayern	1.893097	0.389384	0.596685	0.207301	LK	7.78	
299	9675	Kitzingen	9.0	Bayern	2.451572	0.431702	0.615844	0.184142	LK	6.31	
298	9674	Haßberge	9.0	Bayern	4.570732	0.427365	0.608276	0.180911	LK	6.39	
297	9673	Rhön-Grabfeld	9.0	Bayern	2.391247	0.411238	0.594334	0.183096	LK	5.34	
296	9672	Bad Kissingen	9.0	Bayern	2.362749	0.417143	0.600770	0.183627	LK	6.19	
317	9780	Oberallgäu	9.0	Bayern	-1.079045	0.418779	0.643098	0.224319	LK	6.69	
324	11000	Berlin, Stadt	11.0	Berlin	1.984057	0.467439	0.606226	0.138787	KS	13.12	
328	12054	Potsdam, Stadt	12.0	Brandenburg	3.676391	0.496164	0.638592	0.142428	KS	9.24	
341	12072	Teltow-Fläming	12.0	Brandenburg	10.767790	0.444093	0.600117	0.156024	LK	10.81	
349	13075	Vorpommern-Greifswald	13.0	Mecklenburg-Vorpommern	13.905354	0.482028	0.576674	0.094646	LK	9.78	
348	13074	Nordwestmecklenburg	13.0	Mecklenburg-Vorpommern	9.001209	0.487198	0.603488	0.116290	LK	10.20	
347	13073	Vorpommern-Rügen	13.0	Mecklenburg-Vorpommern	11.521637	0.451591	0.555330	0.103739	LK	10.72	
346	13072	Landkreis Rostock	13.0	Mecklenburg-Vorpommern	10.499634	0.482094	0.611987	0.129892	LK	10.05	
345	13071	Mecklenburgische Seenplatte	13.0	Mecklenburg-Vorpommern	13.431986	0.459977	0.551307	0.091329	LK	10.50	
344	13004	Schwerin	13.0	Mecklenburg-Vorpommern	8.565292	0.451141	0.590019	0.138878	KS	13.42	
343	13003	Rostock	13.0	Mecklenburg-Vorpommern	4.653403	0.409971	0.600771	0.190800	KS	11.46	
342	12073	Uckermark	12.0	Brandenburg	14.394072	0.437368	0.544653	0.107286	LK	9.57	
339	12070	Prignitz	12.0	Brandenburg	13.128358	0.424420	0.558120	0.133700	LK	10.52	
329	12060	Barnim	12.0	Brandenburg	11.378000	0.451209	0.586301	0.135093	LK	9.84	
338	12069	Potsdam-Mittelmark	12.0	Brandenburg	6.867383	0.531080	0.658690	0.127610	LK	7.90	
337	12068	Ostprignitz-Ruppin	12.0	Brandenburg	11.319617	0.421525	0.534547	0.113022	LK	11.79	
336	12067	Oder-Spree	12.0	Brandenburg	12.436741	0.491321	0.591243	0.099923	LK	9.44	
334	12065	Oberhavel	12.0	Brandenburg	9.423493	0.471539	0.611686	0.140147	LK	10.12	
333	12064	Märkisch-Oderland	12.0	Brandenburg	12.269040	0.456014	0.595142	0.139128	LK	9.88	
332	12063	Havelland	12.0	Brandenburg	8.472181	0.460061	0.595300	0.135239	LK	10.33	
331	12062	Elbe-Elster	12.0	Brandenburg	17.468496	0.507831	0.601596	0.093765	LK	8.51	
330	12061	Dahme-Spreewald	12.0	Brandenburg	11.683942	0.512384	0.630916	0.118532	LK	9.52	
214	8416	Tübingen	8.0	Baden-Württemberg	-0.042750	0.577310	0.698302	0.120992	LK	5.52	
200	8235	Calw	8.0	Baden-Württemberg	3.124034	0.518258	0.634144	0.115887	LK	8.02	
212	8337	Waldshut	8.0	Baden-Württemberg	2.672733	0.517815	0.616729	0.098914	LK	8.16	
66	5114	Krefeld, Stadt	5.0	Nordrhein-Westfalen	1.921189	0.483142	0.579292	0.096150	KS	14.48	
76	5166	Viersen	5.0	Nordrhein-Westfalen	1.844257	0.529101	0.620617	0.091516	K	10.41	
75	5162	Rhein-Kreis Neuss	5.0	Nordrhein-Westfalen	2.299495	0.525470	0.630833	0.105362	K	10.52	
74	5158	Mettmann	5.0	Nordrhein-Westfalen	2.260555	0.543117	0.638219	0.095102	K	9.89	
73	5154	Kleve	5.0	Nordrhein-Westfalen	1.841317	0.525321	0.602889	0.077568	K	10.21	
72	5124	Wuppertal, Stadt	5.0	Nordrhein-Westfalen	3.743339	0.481215	0.588017	0.106803	KS	17.89	
71	5122	Solingen, Klingenstadt	5.0	Nordrhein-Westfalen	3.253057	0.469308	0.573856	0.104548	KS	13.53	

	Nr	region	subregion	state	vot19_14	turnout14	turnout19	turnout19_14	state_abbrev	debt_2013	de
70	5120	Remscheid, Stadt	5.0	Nordrhein-Westfalen	3.974041	0.458427	0.563601	0.105175	KS	13.51	
68	5117	Mülheim an der Ruhr, Stadt	5.0	Nordrhein-Westfalen	2.398675	0.523841	0.626328	0.102487	KS	10.61	
67	5116	Mönchengladbach, Stadt	5.0	Nordrhein-Westfalen	3.548520	0.451435	0.548222	0.096787	KS	15.81	
65	5113	Essen, Stadt	5.0	Nordrhein-Westfalen	4.786574	0.473823	0.591842	0.118019	KS	12.80	
49	3451	Ammerland	3.0	Niedersachsen	0.655735	0.492474	0.648330	0.155856	LK	8.84	
63	5111	Düsseldorf, Stadt	5.0	Nordrhein-Westfalen	1.245575	0.538416	0.634823	0.096406	KS	12.47	
58	3460	Vechta	3.0	Niedersachsen	2.555934	0.515049	0.643477	0.128428	LK	7.94	
57	3459	Osnabrück	3.0	Niedersachsen	2.022159	0.529481	0.626696	0.097214	LK	8.63	
56	3458	Oldenburg	3.0	Niedersachsen	1.631730	0.550098	0.629214	0.079116	LK	8.92	
55	3457	Leer	3.0	Niedersachsen	3.745500	0.489421	0.557876	0.068455	LK	10.67	
54	3456	Grafschaft Bentheim	3.0	Niedersachsen	1.965304	0.510500	0.671301	0.160801	LK	8.42	
53	3455	Friesland	3.0	Niedersachsen	1.762954	0.491948	0.620724	0.128776	LK	10.04	
52	3454	Emsland	3.0	Niedersachsen	2.369881	0.513925	0.628178	0.114253	LK	8.35	
51	3453	Cloppenburg	3.0	Niedersachsen	3.756796	0.506470	0.550193	0.043723	LK	9.11	
77	5170	Wesel	5.0	Nordrhein-Westfalen	3.810215	0.527408	0.612658	0.085250	K	9.50	
79	5315	Köln, Stadt	5.0	Nordrhein-Westfalen	0.655069	0.531831	0.646348	0.114517	KS	11.80	
80	5316	Leverkusen, Stadt	5.0	Nordrhein-Westfalen	2.770999	0.488064	0.600015	0.111952	KS	11.15	
81	5334	Städteregion Aachen	5.0	Nordrhein-Westfalen	2.891209	0.545318	0.614457	0.069138	SV	11.02	
101	5766	Lippe	5.0	Nordrhein-Westfalen	4.152088	0.539315	0.612852	0.073536	K	9.56	
100	5762	Höxter	5.0	Nordrhein-Westfalen	3.694157	0.580059	0.622379	0.042320	K	8.03	
99	5758	Herford	5.0	Nordrhein-Westfalen	4.188082	0.527910	0.592793	0.064883	K	10.62	
98	5754	Gütersloh	5.0	Nordrhein-Westfalen	2.861320	0.539049	0.623341	0.084292	K	8.89	
97	5711	Bielefeld, Stadt	5.0	Nordrhein-Westfalen	2.666558	0.533101	0.637928	0.104827	KS	11.50	
96	5570	Warendorf	5.0	Nordrhein-Westfalen	2.709944	0.573719	0.649829	0.076110	K	8.74	
95	5566	Steinfurt	5.0	Nordrhein-Westfalen	2.357753	0.576436	0.649418	0.072981	K	8.85	
94	5562	Recklinghausen	5.0	Nordrhein-Westfalen	5.902598	0.493033	0.583251	0.090218	K	12.18	
93	5558	Coesfeld	5.0	Nordrhein-Westfalen	1.743710	0.601581	0.693412	0.091831	K	7.40	
91	5515	Münster, Stadt	5.0	Nordrhein-Westfalen	-0.118364	0.618601	0.736531	0.117930	KS	8.53	
90	5513	Gelsenkirchen, Stadt	5.0	Nordrhein-Westfalen	8.795457	0.452113	0.512524	0.060411	KS	16.23	
89	5512	Bottrop, Stadt	5.0	Nordrhein-Westfalen	6.703420	0.497681	0.594302	0.096621	KS	11.09	
88	5382	Rhein-Sieg-Kreis	5.0	Nordrhein-Westfalen	2.271056	0.582665	0.659197	0.076532	K	8.96	
87	5378	Rheinisch-Bergischer Kreis	5.0	Nordrhein-Westfalen	1.074745	0.577524	0.680092	0.102567	K	8.77	
86	5374	Oberbergischer Kreis	5.0	Nordrhein-Westfalen	3.539610	0.543416	0.610115	0.066699	K	10.00	
85	5370	Heinsberg	5.0	Nordrhein-Westfalen	4.022322	0.542500	0.589030	0.046530	K	11.29	
84	5366	Euskirchen	5.0	Nordrhein-Westfalen	3.968619	0.529442	0.612848	0.083406	K	10.86	
83	5362	Rhein-Erft-Kreis	5.0	Nordrhein-Westfalen	4.198230	0.522865	0.638116	0.115250	K	10.84	

	Nr	region	subregion	state	vot19_14	turnout14	turnout19	turnout19_14	state_abbrev	debt_2013	de
82	5358	Düren	5.0	Nordrhein-Westfalen	4.444348	0.537839	0.603569	0.065730	K	11.53	
50	3492	Aurich	3.0	Niedersachsen	4.869364	0.507973	0.602403	0.094430	LK	10.00	
48	3405	Wilhelmshaven, Stadt	3.0	Niedersachsen	2.308392	0.375954	0.533773	0.157819	KS	16.22	
104	5911	Bochum, Stadt	5.0	Nordrhein-Westfalen	3.786644	0.501227	0.611073	0.109846	KS	11.97	
12	1060	Segeberg	1.0	Schleswig-Holstein	0.948635	0.423453	0.595325	0.171872	K	10.11	
22	3155	Northeim	3.0	Niedersachsen	2.628395	0.481927	0.560951	0.079024	LK	10.80	
21	3154	Helmstedt	3.0	Niedersachsen	3.786081	0.423608	0.584215	0.160607	LK	10.97	
20	3153	Goslar	3.0	Niedersachsen	3.475091	0.383769	0.584259	0.200491	LK	12.74	
19	3151	Gifhorn	3.0	Niedersachsen	4.910447	0.477766	0.607680	0.129914	LK	9.02	
17	3102	Salzgitter, Stadt	3.0	Niedersachsen	9.198355	0.459825	0.511638	0.051813	KS	12.16	
16	3101	Braunschweig, Stadt	3.0	Niedersachsen	1.221291	0.512913	0.641737	0.128824	KS	10.62	
15	2000	Hamburg, Freie und Hansestadt	2.0	Hamburg	0.472977	0.435025	0.618714	0.183690	KS	10.92	
14	1062	Stormarn	1.0	Schleswig-Holstein	0.523026	0.503786	0.654266	0.150481	K	7.99	
13	1061	Steinburg	1.0	Schleswig-Holstein	1.952564	0.420752	0.577123	0.156371	K	11.43	
11	1059	Schleswig-Flensburg	1.0	Schleswig-Holstein	-0.508833	0.421130	0.598250	0.177120	K	10.45	
47	3404	Osnabrück, Stadt	3.0	Niedersachsen	0.649443	0.509606	0.640557	0.130952	KS	11.28	
10	1058	Rendsburg-Eckernförde	1.0	Schleswig-Holstein	0.028835	0.459591	0.624910	0.165319	K	9.16	
9	1057	Plön	1.0	Schleswig-Holstein	0.307204	0.469130	0.636865	0.167735	K	8.67	
8	1056	Pinneberg	1.0	Schleswig-Holstein	0.703336	0.457528	0.629707	0.172180	K	9.67	
7	1055	Ostholstein	1.0	Schleswig-Holstein	0.252615	0.424788	0.584264	0.159476	K	10.63	
6	1054	Nordfriesland	1.0	Schleswig-Holstein	-0.013827	0.411905	0.588692	0.176788	K	10.09	
5	1053	Herzogtum Lauenburg	1.0	Schleswig-Holstein	1.646033	0.463842	0.601961	0.138119	K	10.16	
4	1051	Dithmarschen	1.0	Schleswig-Holstein	3.374651	0.397193	0.544108	0.146915	K	12.52	
3	1004	Neumünster, Stadt	1.0	Schleswig-Holstein	2.392481	0.453649	0.482205	0.028556	KS	16.61	
2	1003	Lübeck, Hansestadt	1.0	Schleswig-Holstein	0.551817	0.376398	0.546124	0.169726	KS	15.25	
23	3157	Peine	3.0	Niedersachsen	4.032544	0.468641	0.619507	0.150865	LK	9.78	
24	3158	Wolfenbüttel	3.0	Niedersachsen	2.677336	0.525676	0.644388	0.118711	LK	9.13	
26	3241	Region Hannover	3.0	Niedersachsen	2.288645	0.486508	0.639614	0.153106	LK	12.01	
27	3251	Diepholz	3.0	Niedersachsen	2.283782	0.495989	0.621981	0.125992	LK	9.23	
46	3403	Oldenburg (Oldenburg), Stadt	3.0	Niedersachsen	-0.349140	0.474169	0.648877	0.174708	KS	11.25	
45	3402	Emden, Stadt	3.0	Niedersachsen	4.692135	0.359377	0.603340	0.243963	KS	13.57	
44	3401	Delmenhorst, Stadt	3.0	Niedersachsen	5.378949	0.445939	0.526147	0.080208	KS	15.53	
43	3361	Verden	3.0	Niedersachsen	2.500429	0.533703	0.642220	0.108517	LK	8.99	
42	3360	Uelzen	3.0	Niedersachsen	2.842799	0.498988	0.617496	0.118508	LK	11.60	
41	3359	Stade	3.0	Niedersachsen	2.274053	0.479329	0.622718	0.143389	LK	9.10	
40	3358	Heidekreis	3.0	Niedersachsen	3.528682	0.510705	0.592424	0.081719	LK	10.84	
39	3357	Rotenburg (Wümme)	3.0	Niedersachsen	3.029319	0.554893	0.621731	0.066838	LK	9.44	
38	3356	Osterholz	3.0	Niedersachsen	2.277387	0.497100	0.626313	0.129213	LK	8.78	
37	3355	Lüneburg	3.0	Niedersachsen	1.548526	0.531335	0.669420	0.138085	LK	9.97	
36	3354	Lüchow-Dannenberg	3.0	Niedersachsen	2.486280	0.529620	0.589581	0.059961	LK	11.55	
35	3353	Harburg	3.0	Niedersachsen	-0.329070	0.522290	0.664795	0.142504	LK	8.12	

	Nr	region	subregion	state	vot19_14	turnout14	turnout19	turnout19_14	state_abbrev	debt_2013	de
34	3352	Cuxhaven	3.0	Niedersachsen	3.252164	0.461184	0.618225	0.157041	LK	10.88	
33	3351	Celle	3.0	Niedersachsen	2.984502	0.465224	0.601866	0.136642	LK	11.90	
32	3257	Schaumburg	3.0	Niedersachsen	3.185247	0.506238	0.586828	0.080590	LK	11.54	
31	3256	Nienburg (Weser)	3.0	Niedersachsen	3.977201	0.431191	0.567161	0.135970	LK	10.26	
30	3255	Holzminden	3.0	Niedersachsen	3.587561	0.470233	0.598803	0.128569	LK	11.91	
29	3254	Hildesheim	3.0	Niedersachsen	2.812249	0.484514	0.624283	0.139769	LK	10.98	
28	3252	Hameln-Pyrmont	3.0	Niedersachsen	3.762241	0.491912	0.584282	0.092370	LK	12.23	
102	5770	Minden-Lübbecke	5.0	Nordrhein-Westfalen	3.816122	0.508022	0.590705	0.082683	K	10.00	
105	5913	Dortmund, Stadt	5.0	Nordrhein-Westfalen	3.246025	0.474730	0.584610	0.109879	KS	14.01	
211	8336	Lörrach	8.0	Baden-Württemberg	2.124510	0.485008	0.603621	0.118613	LK	8.30	
175	7338	Rhein-Pfalz-Kreis	7.0	Rheinland-Pfalz	3.608337	0.595330	0.689203	0.093873	LK	7.95	
185	8125	Heilbronn	8.0	Baden-Württemberg	4.557445	0.517770	0.640056	0.122286	LK	6.79	
184	8121	Heilbronn, Stadt	8.0	Baden-Württemberg	4.498336	0.424881	0.554965	0.130084	SK	10.51	
183	8119	Rems-Murr-Kreis	8.0	Baden-Württemberg	1.459988	0.528989	0.649230	0.120241	LK	7.87	
182	8118	Ludwigsburg	8.0	Baden-Württemberg	1.606565	0.545910	0.668405	0.122494	LK	7.68	
181	8117	Göppingen	8.0	Baden-Württemberg	3.906186	0.502929	0.626970	0.124041	LK	7.91	
179	8115	Böblingen	8.0	Baden-Württemberg	1.319050	0.536564	0.661765	0.125201	LK	6.72	
178	8111	Stuttgart, Landeshauptstadt	8.0	Baden-Württemberg	-0.004270	0.531539	0.670551	0.139012	SK	10.72	
177	7340	Südwestpfalz	7.0	Rheinland-Pfalz	6.540424	0.668604	0.711160	0.042556	LK	7.91	
176	7339	Mainz-Bingen	7.0	Rheinland-Pfalz	2.114120	0.626516	0.707297	0.080782	LK	8.28	
174	7337	Südliche Weinstraße	7.0	Rheinland-Pfalz	3.677925	0.632733	0.714668	0.081935	LK	7.93	
160	7313	Landau in der Pfalz, kreisfreie Stadt	7.0	Rheinland-Pfalz	0.609086	0.546693	0.665361	0.118668	KS	9.54	
173	7336	Kusel	7.0	Rheinland-Pfalz	7.428792	0.613561	0.666013	0.052452	LK	9.42	
171	7334	Germersheim	7.0	Rheinland-Pfalz	6.502230	0.578296	0.657366	0.079070	LK	8.73	
170	7333	Donnersbergkreis	7.0	Rheinland-Pfalz	5.454532	0.592133	0.664093	0.071960	LK	10.87	
169	7332	Bad Dürkheim	7.0	Rheinland-Pfalz	4.193089	0.603350	0.699289	0.095939	LK	9.06	
168	7331	Alzey-Worms	7.0	Rheinland-Pfalz	4.706680	0.609384	0.678552	0.069168	LK	9.65	
167	7320	Zweibrücken, kreisfreie Stadt	7.0	Rheinland-Pfalz	4.552377	0.436552	0.527215	0.090663	KS	12.01	
166	7319	Worms, kreisfreie Stadt	7.0	Rheinland-Pfalz	4.672871	0.488433	0.554494	0.066061	KS	15.52	
163	7316	Neustadt an der Weinstraße, kreisfreie Stadt	7.0	Rheinland-Pfalz	3.732054	0.551242	0.653283	0.102041	KS	10.56	
162	7315	Mainz, kreisfreie Stadt	7.0	Rheinland-Pfalz	-0.124031	0.550887	0.681626	0.130739	KS	8.16	
186	8126	Hohenlohekreis	8.0	Baden-Württemberg	4.833844	0.514650	0.628298	0.113647	LK	5.88	
187	8127	Schwäbisch Hall	8.0	Baden-Württemberg	4.244568	0.481910	0.606205	0.124295	LK	6.60	
188	8128	Main-Tauber-Kreis	8.0	Baden-Württemberg	3.981060	0.574151	0.662187	0.088036	LK	5.76	
189	8135	Heidenheim	8.0	Baden-Württemberg	2.764367	0.468331	0.583296	0.114965	LK	7.63	

	Nr	region	subregion	state	vot19_14	turnout14	turnout19	turnout19_14	state_abbrev	debt_2013	de
210	8335	Konstanz	8.0	Baden-Württemberg	1.091857	0.506818	0.633380	0.126562	LK	10.00	
209	8327	Tuttlingen	8.0	Baden-Württemberg	3.670103	0.506034	0.606472	0.100438	LK	8.55	
208	8326	Schwarzwald-Baar-Kreis	8.0	Baden-Württemberg	1.631454	0.492172	0.585412	0.093241	LK	8.81	
207	8325	Rottweil	8.0	Baden-Württemberg	2.742601	0.515141	0.617389	0.102248	LK	7.11	
206	8317	Ortenaukreis	8.0	Baden-Württemberg	2.870258	0.509676	0.608586	0.098910	LK	7.91	
204	8315	Breisgau-Hochschwarzwald	8.0	Baden-Württemberg	0.179041	0.574263	0.673257	0.098994	LK	7.02	
203	8311	Freiburg im Breisgau, Stadt	8.0	Baden-Württemberg	-0.395556	0.550328	0.673976	0.123648	SK	7.92	
202	8237	Freudenstadt	8.0	Baden-Württemberg	4.655189	0.511898	0.614363	0.102466	LK	6.92	
201	8236	Enzkreis	8.0	Baden-Württemberg	1.503058	0.536281	0.656299	0.120018	LK	6.69	
199	8231	Pforzheim, Stadt	8.0	Baden-Württemberg	3.168329	0.381547	0.532481	0.150934	SK	13.19	
198	8226	Rhein-Neckar-Kreis	8.0	Baden-Württemberg	2.000629	0.541816	0.658812	0.116996	LK	7.90	
197	8225	Neckar-Odenwald-Kreis	8.0	Baden-Württemberg	5.161887	0.542641	0.636719	0.094077	LK	7.40	
196	8222	Mannheim, Universitätsstadt	8.0	Baden-Württemberg	0.973703	0.440724	0.583170	0.142446	SK	13.36	
195	8221	Heidelberg, Stadt	8.0	Baden-Württemberg	-0.606952	0.547706	0.701307	0.153601	SK	6.29	
194	8216	Rastatt	8.0	Baden-Württemberg	2.820656	0.498231	0.620487	0.122257	LK	7.06	
193	8215	Karlsruhe	8.0	Baden-Württemberg	2.761304	0.536451	0.649101	0.112650	LK	7.12	
192	8212	Karlsruhe, Stadt	8.0	Baden-Württemberg	0.102003	0.489247	0.646261	0.157014	SK	8.70	
191	8211	Baden-Baden, Stadt	8.0	Baden-Württemberg	0.493269	0.475253	0.616598	0.141345	SK	8.89	
190	8136	Ostalbkreis	8.0	Baden-Württemberg	3.027529	0.503761	0.629620	0.125860	LK	6.85	
161	7314	Ludwigshafen am Rhein, kreisfreie Stadt	7.0	Rheinland-Pfalz	4.514345	0.450383	0.540737	0.090354	KS	15.26	
157	7235	Trier-Saarburg	7.0	Rheinland-Pfalz	1.341529	0.615229	0.701806	0.086577	LK	6.76	
106	5914	Hagen, Stadt der FernUniversität	5.0	Nordrhein-Westfalen	6.283618	0.480564	0.556763	0.076199	KS	14.92	
118	6413	Offenbach am Main, Stadt	6.0	Hessen	0.554018	0.371397	0.515239	0.143842	KS	18.61	
128	6439	Rheingau-Taunus-Kreis	6.0	Hessen	0.160638	0.463464	0.646824	0.183360	LK	7.77	
127	6438	Offenbach	6.0	Hessen	0.501197	0.431652	0.598019	0.166367	LK	9.06	
126	6437	Odenwaldkreis	6.0	Hessen	2.279088	0.433513	0.563363	0.129851	LK	9.55	
124	6435	Main-Kinzig-Kreis	6.0	Hessen	1.851829	0.399348	0.562324	0.162976	LK	9.85	
123	6434	Hochtaunuskreis	6.0	Hessen	-2.160846	0.504762	0.651337	0.146576	LK	7.07	
122	6433	Groß-Gerau	6.0	Hessen	1.373835	0.420806	0.572989	0.152183	LK	9.29	
121	6432	Darmstadt-Dieburg	6.0	Hessen	1.006714	0.455993	0.633141	0.177149	LK	8.31	
120	6431	Bergstraße	6.0	Hessen	2.373800	0.453391	0.605093	0.151702	LK	8.84	
119	6414	Wiesbaden, Landeshauptstadt	6.0	Hessen	-0.386746	0.414751	0.590462	0.175712	KS	16.18	
117	6412	Frankfurt am Main, Stadt	6.0	Hessen	-1.225643	0.445161	0.600908	0.155747	KS	11.04	
156	7233	Vulkaneifel	7.0	Rheinland-Pfalz	1.838407	0.605811	0.640866	0.035054	LK	8.32	
115	5978	Unna	5.0	Nordrhein-Westfalen	4.389196	0.512944	0.596300	0.083356	K	11.26	

	Nr	region	subregion	state	vot19_14	turnout14	turnout19	turnout19_14	state_abbrev	debt_2013	de
114	5974	Soest	5.0	Nordrhein-Westfalen	3.174263	0.530933	0.604598	0.073665	K	10.30	
113	5970	Siegen-Wittgenstein	5.0	Nordrhein-Westfalen	4.081969	0.534922	0.615293	0.080370	K	9.10	
112	5966	Olpe	5.0	Nordrhein-Westfalen	1.432137	0.541969	0.634655	0.092686	K	8.91	
111	5962	Märkischer Kreis	5.0	Nordrhein-Westfalen	3.556639	0.468920	0.563104	0.094184	K	12.40	
110	5958	Hochsauerlandkreis	5.0	Nordrhein-Westfalen	2.847183	0.553803	0.620682	0.066880	K	10.37	
109	5954	Ennepe-Ruhr-Kreis	5.0	Nordrhein-Westfalen	3.252381	0.522540	0.620442	0.097902	K	10.66	
108	5916	Herne, Stadt	5.0	Nordrhein-Westfalen	6.761899	0.440452	0.536577	0.096125	KS	16.04	
107	5915	Hamm, Stadt	5.0	Nordrhein-Westfalen	6.432973	0.531568	0.577385	0.045817	KS	14.15	
129	6440	Wetteraukreis	6.0	Hessen	0.576312	0.402559	0.584662	0.182103	LK	9.75	
130	6531	Gießen	6.0	Hessen	1.061274	0.410387	0.574934	0.164547	LK	9.23	
132	6533	Limburg-Weilburg	6.0	Hessen	0.915756	0.385065	0.552957	0.167892	LK	10.35	
133	6534	Marburg-Biedenkopf	6.0	Hessen	1.804533	0.392987	0.568983	0.175996	LK	8.00	
155	7232	Eifelkreis Bitburg-Prüm	7.0	Rheinland-Pfalz	2.399063	0.628351	0.693222	0.064871	LK	7.72	
154	7231	Bernkastel-Wittlich	7.0	Rheinland-Pfalz	2.432175	0.595588	0.650677	0.055089	LK	8.00	
153	7211	Trier, kreisfreie Stadt	7.0	Rheinland-Pfalz	0.845781	0.485665	0.603163	0.117498	KS	10.14	
152	7143	Westerwaldkreis	7.0	Rheinland-Pfalz	3.253560	0.588877	0.653080	0.064203	LK	10.77	
150	7140	Rhein-Hunsrück-Kreis	7.0	Rheinland-Pfalz	3.205590	0.617591	0.682011	0.064420	LK	8.18	
149	7138	Neuwied	7.0	Rheinland-Pfalz	1.710973	0.542383	0.618731	0.076348	LK	10.69	
148	7137	Mayen-Koblenz	7.0	Rheinland-Pfalz	2.471077	0.554617	0.622215	0.067598	LK	9.82	
146	7134	Birkenfeld	7.0	Rheinland-Pfalz	4.346890	0.529860	0.585639	0.055780	LK	10.26	
145	7133	Bad Kreuznach	7.0	Rheinland-Pfalz	3.621077	0.585890	0.638462	0.052572	LK	10.11	
144	7132	Altenkirchen (Westerwald)	7.0	Rheinland-Pfalz	3.542119	0.552569	0.608409	0.055840	LK	11.43	
143	7131	Ahrweiler	7.0	Rheinland-Pfalz	0.707541	0.574314	0.659309	0.084994	LK	8.72	
142	7111	Koblenz, kreisfreie Stadt	7.0	Rheinland-Pfalz	0.766647	0.493014	0.607569	0.114556	KS	12.73	
141	6636	Werra-Meißner-Kreis	6.0	Hessen	1.931501	0.422513	0.556307	0.133794	LK	10.92	
140	6635	Waldeck-Frankenberg	6.0	Hessen	2.278377	0.371918	0.522278	0.150360	LK	9.71	
139	6634	Schwalm-Eder-Kreis	6.0	Hessen	2.128785	0.432209	0.560155	0.127945	LK	9.65	
138	6633	Kassel	6.0	Hessen	0.131689	0.413184	0.565107	0.151923	LK	9.06	
137	6632	Hersfeld-Rotenburg	6.0	Hessen	3.386571	0.389272	0.551690	0.162418	LK	9.11	
136	6631	Fulda	6.0	Hessen	3.861435	0.391362	0.571427	0.180065	LK	7.52	
134	6535	Vogelsbergkreis	6.0	Hessen	2.893113	0.417771	0.563673	0.145902	LK	8.41	
400	16077	Altenburger Land	16.0	Thüringen	19.110053	0.483017	0.565661	0.082643	LK	8.29	

- From the describe function we saw mean,median were approximately same , but as we suspected ouliers we will visualize them ,But in order to visualize them uisng box plot we remove nan values from data .
- For nan value **IMPUTATION** use **Median** as there are possible ouliers.
- As the median deals well with ouliers it is a better imputation when compared to mean.

In [3]:

```
df1.fillna(df1.median(),inplace = True)
```

In [4]:

```
df1.isnull().all().sum()
# wehave removed all the nan values from the data now lets visualize the attributes
```

Out[4]:

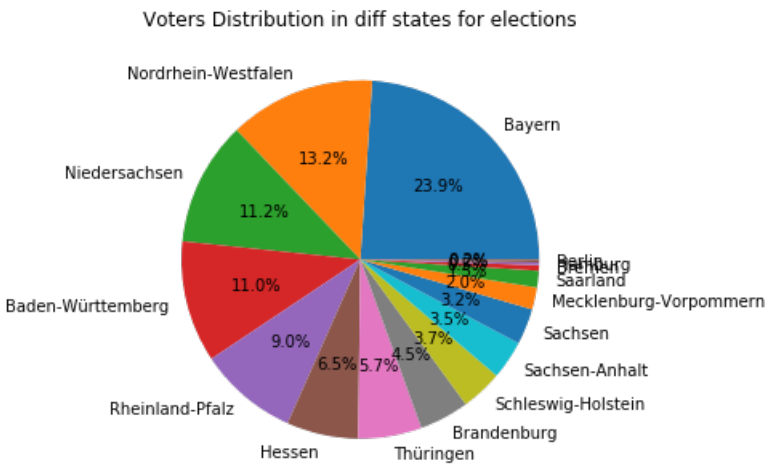
0

Data visualization:

- Brief knowledge of the data.
- Indetifying the target varaible.
- identiying the turnout ratio patterns regarding different attributes.
- Identifying the key attributes that constitute to the target variable.
- uni,Bi, Multivariate analysis.
- knowledge of the distribution inside the data.

In [46]:

```
fig = plt.figure(figsize=(5,10))
df1['state'].value_counts().plot(kind = 'pie', autopct='%.1f%%')
plt.ylabel(" ", fontsize = 25)
plt.title("Voters Distribution in diff states for elections")
print("")
```



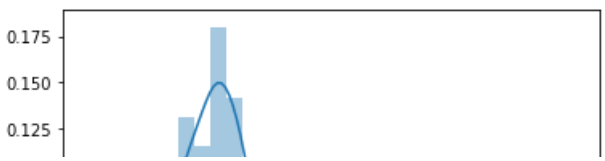
Bayern state people have voted highest for AFD in the election (23.9%) . i.e change rates are higher for this state

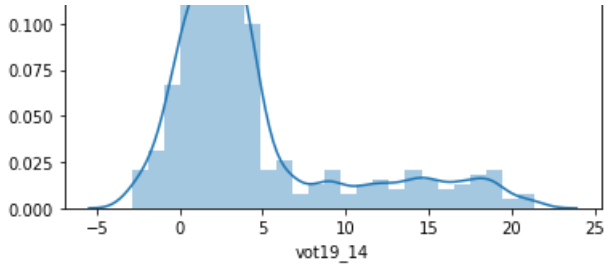
- **Berlin state is the state voted the least for AFD (0.2%).**

In [47]:

```
sns.distplot(df1['vot19_14']);
#skewness and kurtosis
print("Skewness: %f" % df1['vot19_14'].skew())
print("Kurtosis: %f" % df1['vot19_14'].kurt())
```

```
Skewness: 1.511796
Kurtosis: 1.474949
```





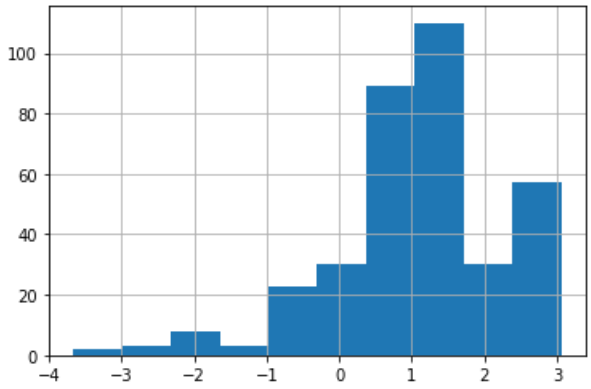
- From the above plot we can say that the target column is **right** skewed as we can clearly see that mean > median. so most of the data point fall towars the positive skew.
- Which is why we have imputed the missing values with median.
- the data is platykurtic less than the meso or normal distribution.

In [9]:

```
df1["vot19_14"].apply(np.log).hist()
```

Out[9]:

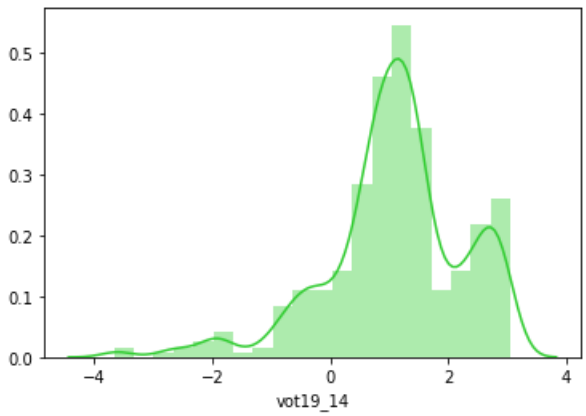
```
<matplotlib.axes._subplots.AxesSubplot at 0x1de2ff30508>
```



In [14]:

```
sns.distplot(np.log(df1.vot19_14), bins=20, color="limegreen");
```

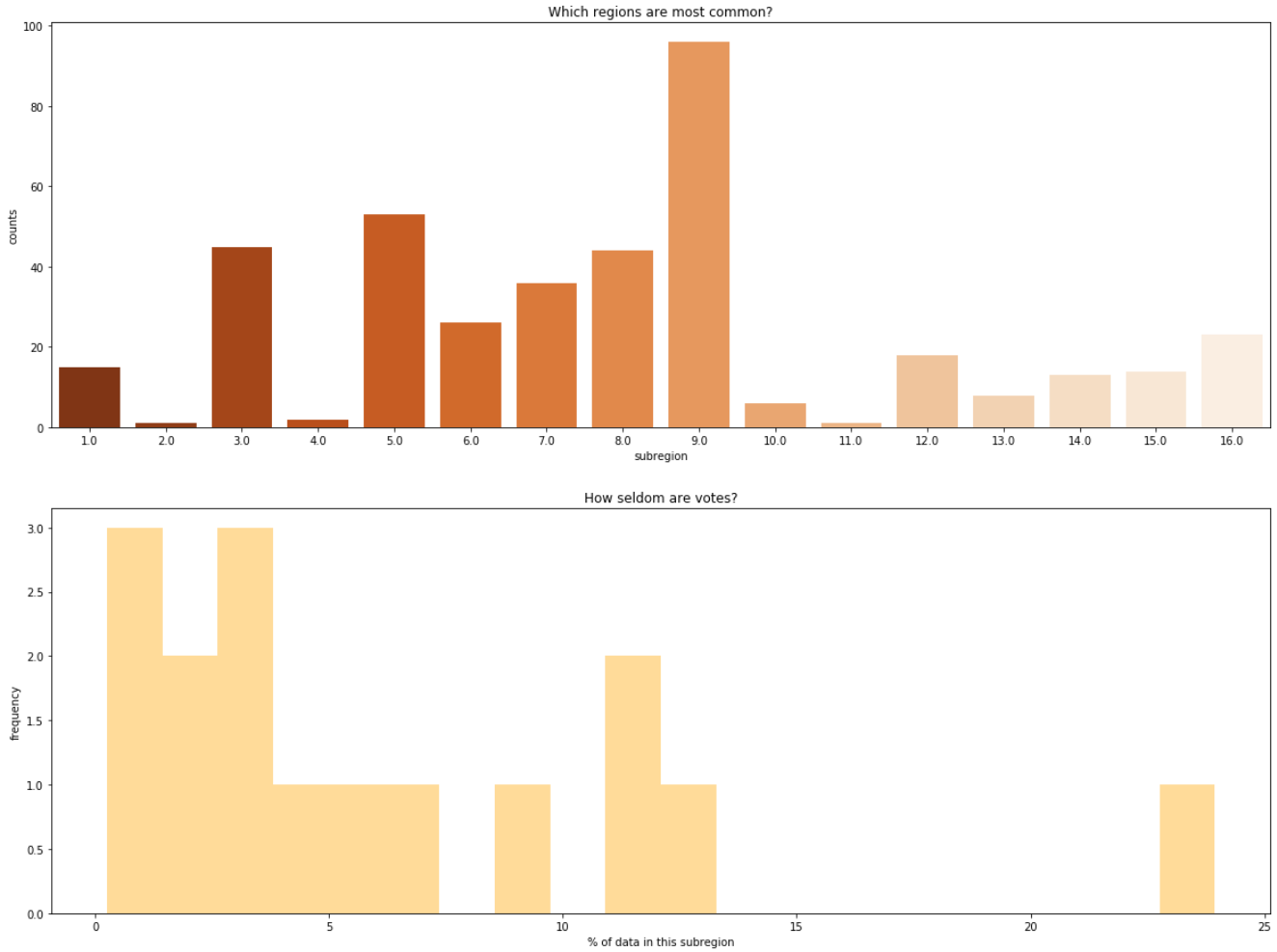
```
C:\Users\SOURAV CH\Anaconda3\lib\site-packages\statsmodels\nonparametric\kde.py:447:
RuntimeWarning: invalid value encountered in greater
  X = X[np.logical_and(X > clip[0], X < clip[1])] # won't work for two columns.
C:\Users\SOURAV CH\Anaconda3\lib\site-packages\statsmodels\nonparametric\kde.py:447:
RuntimeWarning: invalid value encountered in less
  X = X[np.logical_and(X > clip[0], X < clip[1])] # won't work for two columns.
```



Sometimes the target migth be a heavily skewed data no actually know it could or not be normal , simply if we log transfrom it we can see the trends in it.

In [4]:

```
vote_counts = df1.subregion.value_counts().sort_values(ascending=False)
fig, ax = plt.subplots(2,1,figsize=(20,15))
sns.barplot(vote_counts.iloc[0:20].index,
            vote_counts.iloc[0:20].values,
            ax = ax[0], palette="Oranges_r")
ax[0].set_ylabel("counts")
ax[0].set_xlabel("subregion")
ax[0].set_title("Which regions are most common?");
sns.distplot(np.round(vote_counts/df1.shape[0]*100,2),
             kde=False,
             bins=20,
             ax=ax[1], color="Orange")
ax[1].set_title("How seldom are votes?")
ax[1].set_xlabel("% of data in this subregion")
ax[1].set_ylabel("frequency");
```



The above viz gives us the most repeated subregion i.e which sub region actually caasted their votes in 2019-2014, the viz below actually gives the evidence about by showing the shaded region as votes freq of a particular Subregion.

In [12]:

```
pivot = df1.pivot_table(index='debt_2018', values='vot19_14', aggfunc=np.median)
pivot_1 = pivot.head(15)
pivot_1
```

Out[12]:

debt_2018	vot19_14
3.85	2.957736
4.99	0.073017

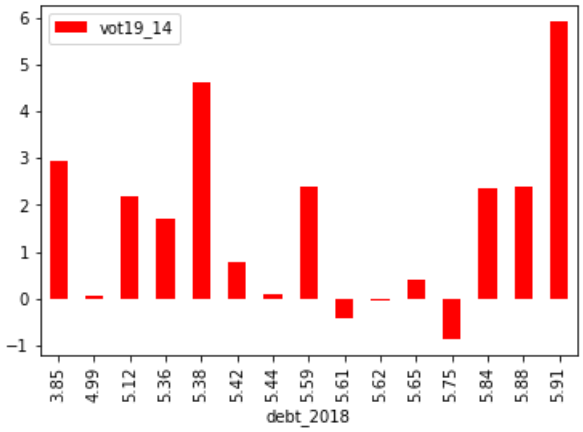
debt_2018	vot19_14
5.12	2.185083
5.36	1.694857
5.38	4.637548
5.42	0.767276
5.44	0.080474
5.59	2.400047
5.61	-0.425402
5.62	-0.057201
5.65	0.388621
5.75	-0.857189
5.84	2.370208
5.88	2.391247
5.91	5.929222

In [23]:

```
pivot_1.plot(kind='bar', color='red') # a simple plot for debt and vot_19
```

Out[23]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1de319f0b48>
```



In [28]:

```
pivot = df1.pivot_table(index='state', values='vot19_14', aggfunc=np.median)
pivot_1 = pivot.head(50)
pivot_1
```

Out[28]:

state	vot19_14
Baden-Württemberg	2.414498
Bayern	0.777401
Berlin	1.984057
Brandenburg	11.530971
Bremen	3.169748
Hamburg	0.472977
Hessen	1.033994
Mecklenburg-Vorpommern	10.896717
Niedersachsen	2.677336
Nordrhein-Westfalen	3.253057
Rheinland-Pfalz	3.575548

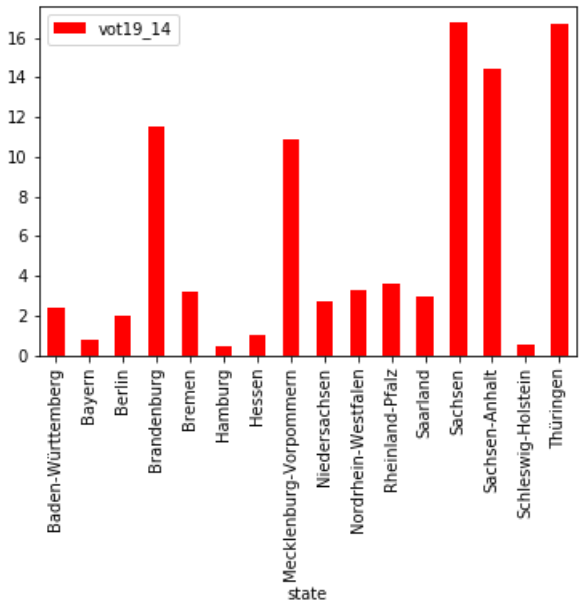
Saarland	2.819504
Sachsen	16.732576
Sachsen-Anhalt	14.438366
Schleswig-Holstein	0.523026
Thüringen	16.664091

In [29]:

```
pivot_1.plot(kind='bar', color='red') #
```

Out[29]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1de31b4dec8>
```



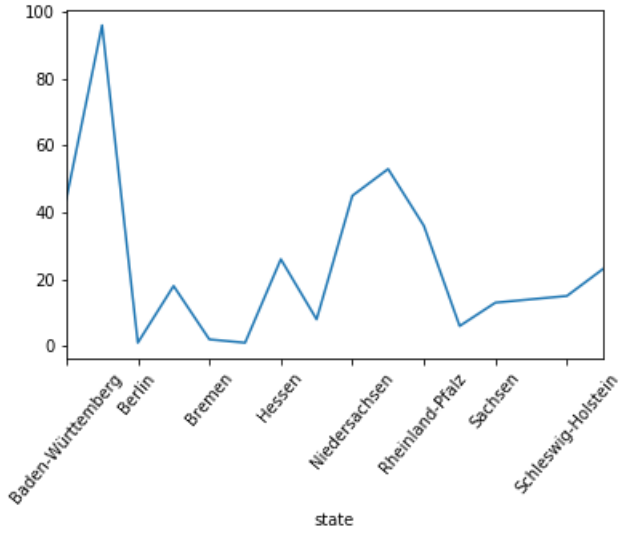
The above plot gives us the information about the percentage of voters in 19_14 election change in different subregions

In [5]:

```
n_by_state = df1.groupby("state")["turnout19"].count().plot()
plt.xticks(rotation=50)
n_by_state
```

Out[5]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x197b7fbbf08>
```



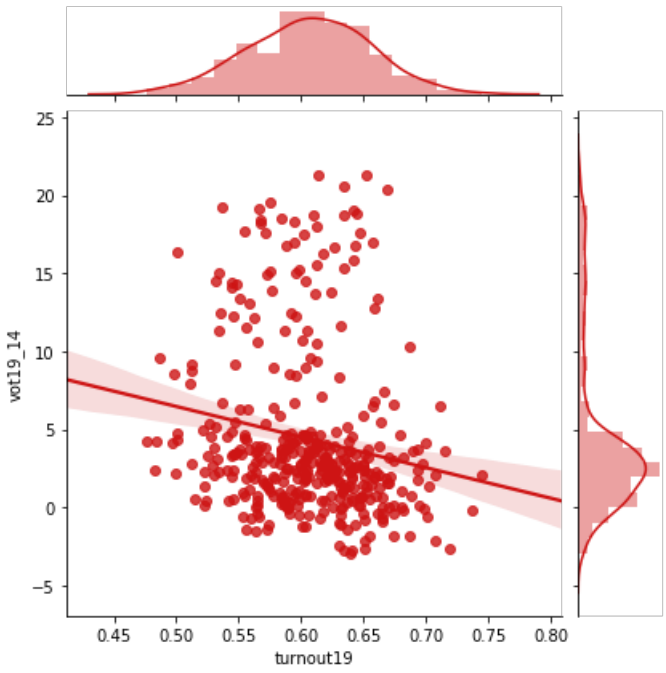
Visualization of the states that actually turned out in 2019 and their count %s

In [13]:

```
sns.jointplot(df1.loc[:,'turnout19'], df1.loc[:,'vot19_14'], kind="regg", color="#ce1414")
```

Out[13]:

```
<seaborn.axisgrid.JointGrid at 0x26671092648>
```



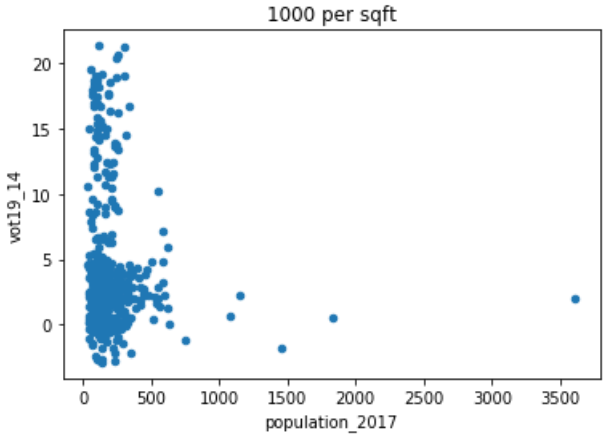
The above plot was to basically understand the percentage of votes of 2019 turnouts in the change over rate , knowing this would actually give us the info about how many valid turnouts were from 2019 in the change over rate.

In [12]:

```
df1.plot.scatter(x='population_2017', y='vot19_14', title='1000 per sqft')
```

Out[12]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x26670058e08>
```

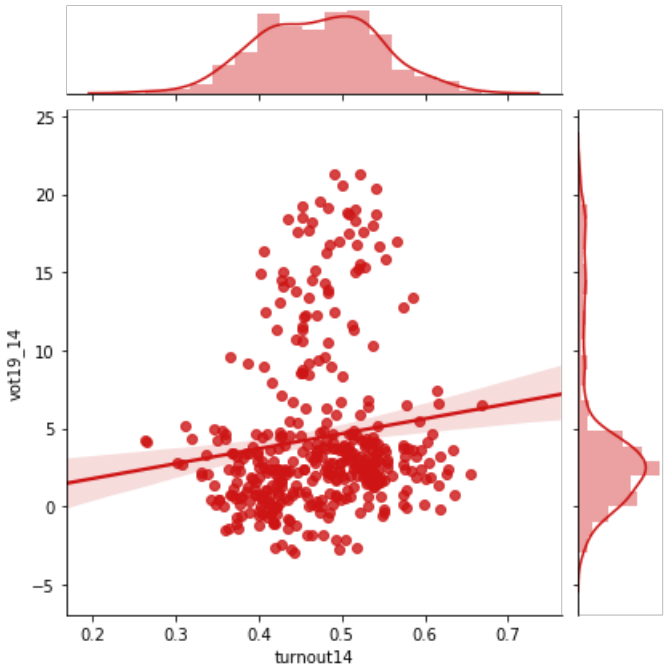


In [30]:

```
sns.jointplot(df1.loc[:,'turnout14'], df1.loc[:,'vot19_14'], kind="regg", color="#ce1414")
```

Out[30]:

```
<seaborn.axisgrid.JointGrid at 0x1de31bfc788>
```



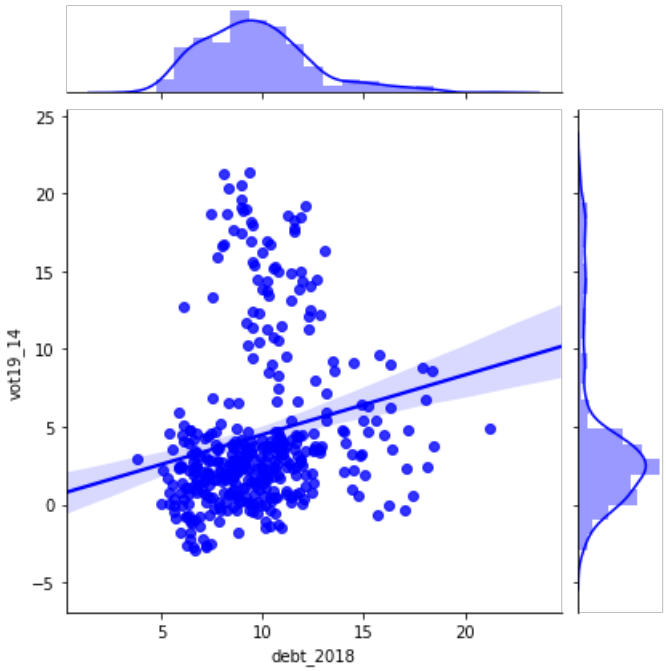
The above plot was to basically understand the percentage of votes of 2014 turnouts in the change over rate , knowing this would actually give us the info about how many valid turnouts were from 2014 in the change over rate.

In [81]:

```
sns.jointplot(df1.loc[:,'debt_2018'], df1.loc[:,'vot19_14'], kind="regg", color="blue")
```

Out[81]:

```
<seaborn.axisgrid.JointGrid at 0x1f4ee306358>
```

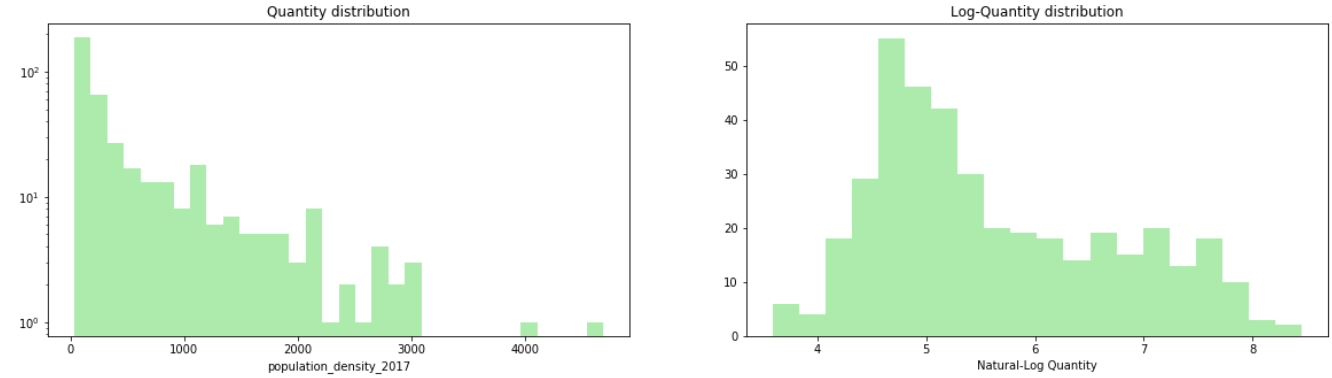


The above distribution plot gives us the info about trends like if debt were around 15% in 2018 what was the change over rate ahving high and low debts.

In [86]:

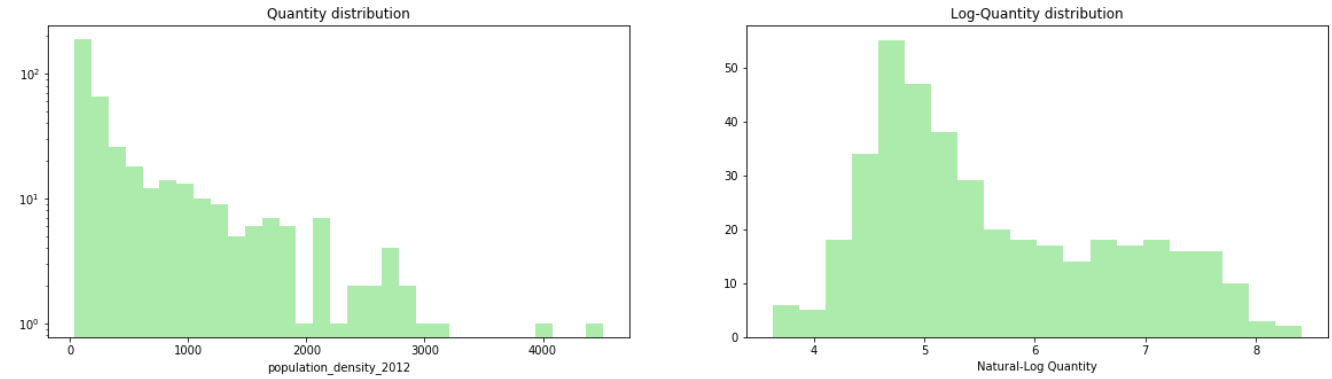
```
fig, ax = plt.subplots(1,2,figsize=(20,5))
sns.distplot(df1.population_density_2017, ax=ax[0], kde=False, color="limegreen");
```

```
sns.distplot(np.log(df1.population_density_2017), ax=ax[1], bins=20, kde=False, color="limegreen");
ax[0].set_title("Quantity distribution")
ax[0].set_yscale("log")
ax[1].set_title("Log-Quantity distribution")
ax[1].set_xlabel("Natural-Log Quantity");

# below plot is to visualize the population density of all the sates collectively
```



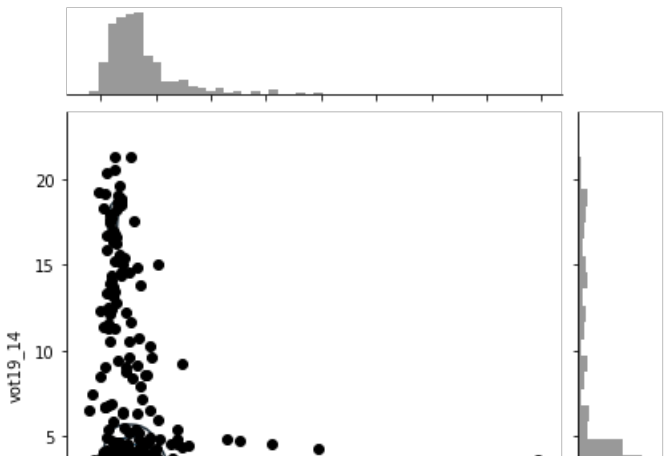
In [17]:

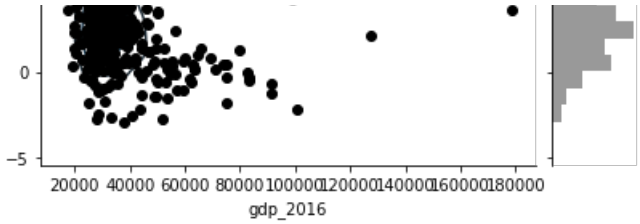
```
fig, ax = plt.subplots(1,2,figsize=(20,5))
sns.distplot(df1.population_density_2012, ax=ax[0], kde=False, color="limegreen");
sns.distplot(np.log(df1.population_density_2012), ax=ax[1], bins=20, kde=False, color="limegreen");
ax[0].set_title("Quantity distribution")
ax[0].set_yscale("log")
ax[1].set_title("Log-Quantity distribution")
ax[1].set_xlabel("Natural-Log Quantity");

# below plot is to visualize the population density of all the sates collectively
```



In [32]:

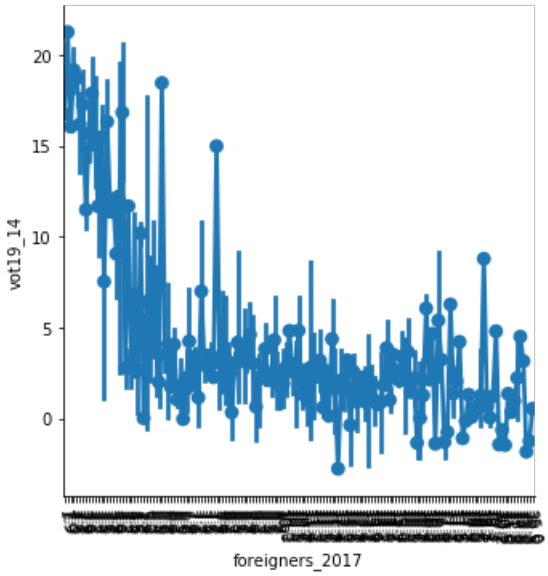
```
g = (sns.jointplot("gdp_2016", "vot19_14",
                   data=df1, color="k")
        .plot_joint(sns.kdeplot, zorder=0, n_levels=6))
```





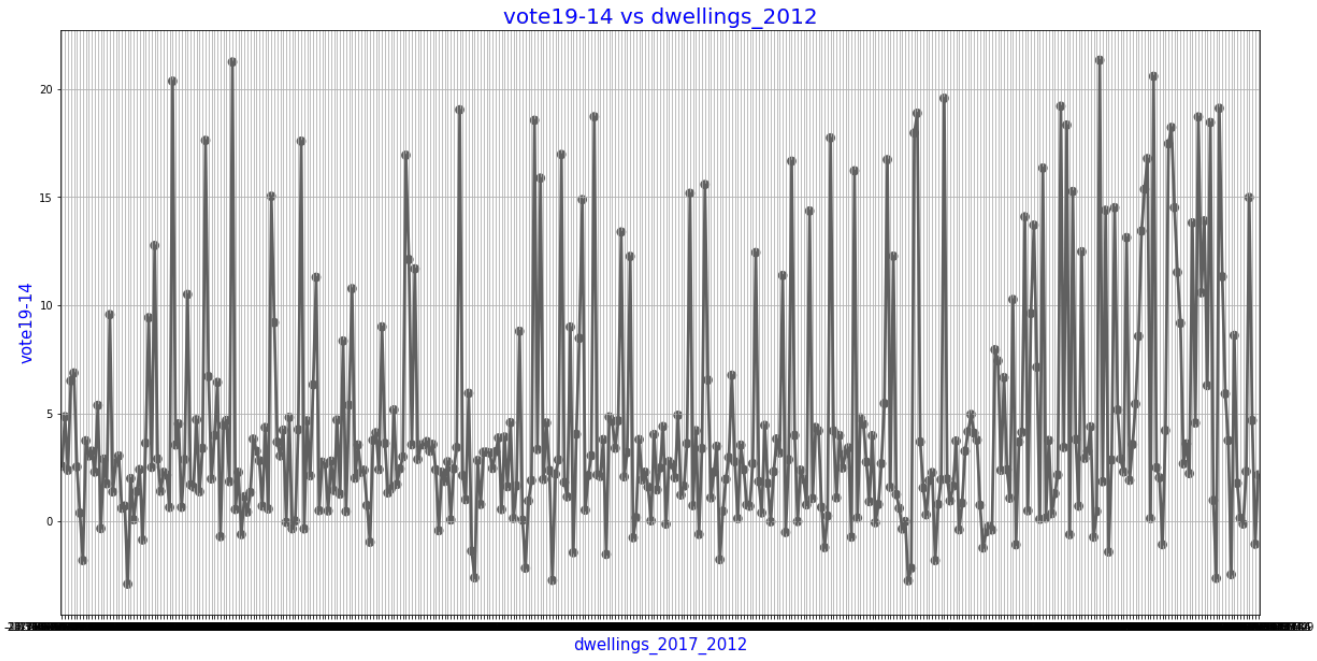
In [114]:

```
sns.factorplot('foreigners_2017','vot19_14',data=df1,samples = 10)
plt.xticks(rotation=1000000000)
plt.show()
```



In [99]:

```
f,ax1 = plt.subplots(figsize =(20,10))
sns.pointplot(x='dwellings_2017_2012',y='vot19_14',data=df1,color='#606060',alpha=0.8)
plt.xlabel('dwellings_2017_2012',fontsize = 15,color='blue')
plt.ylabel('vote19-14',fontsize = 15,color='blue')
plt.title('vote19-14 vs dwellings_2012',fontsize = 20,color='blue')
plt.grid()
plt.show()
```



From the above plot we can infer that as the no of dwellings got increased by 2019 from 2012 , population got increased which might be a major factor for AFd turnout to parliament when compared to 2014

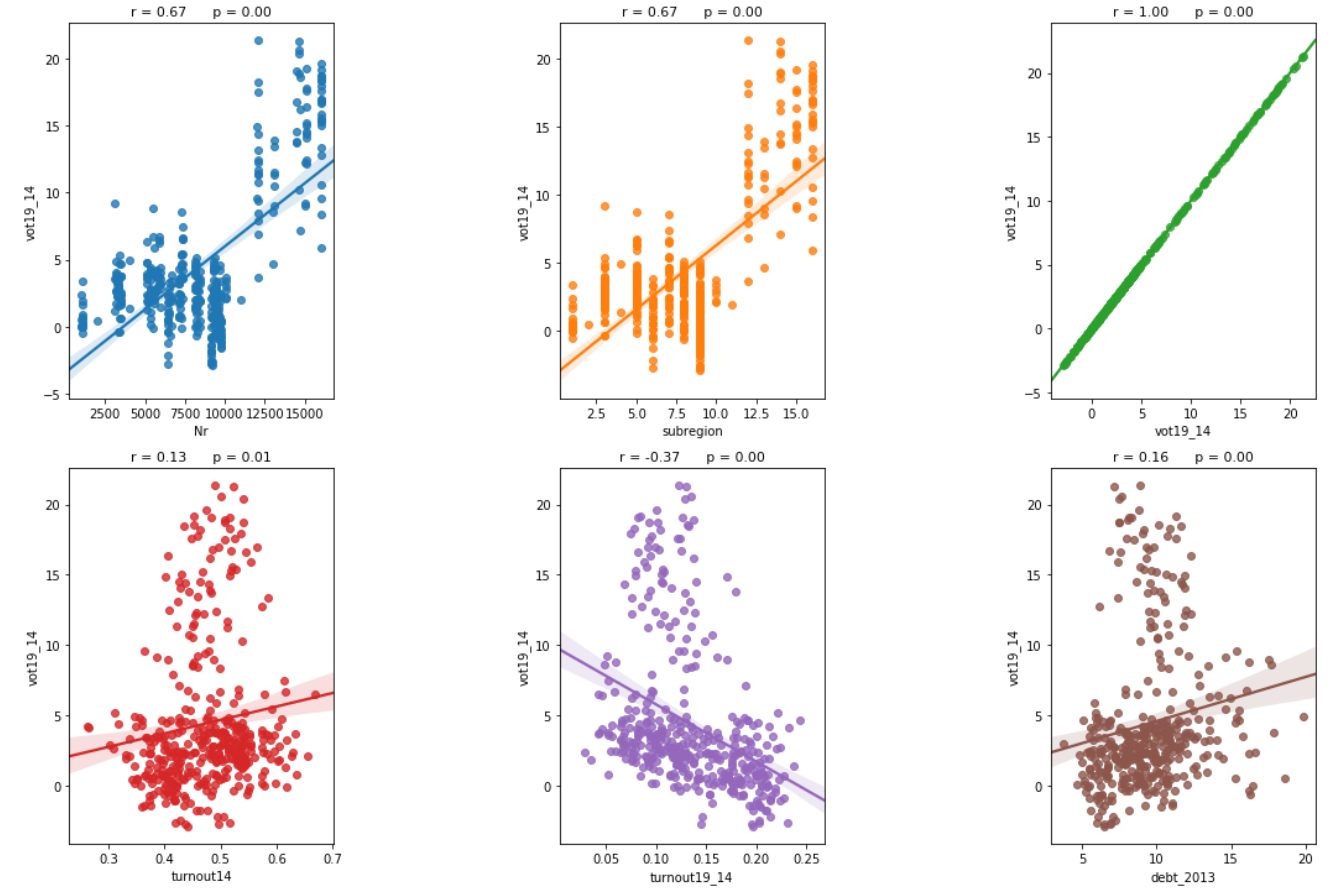
In [5]:

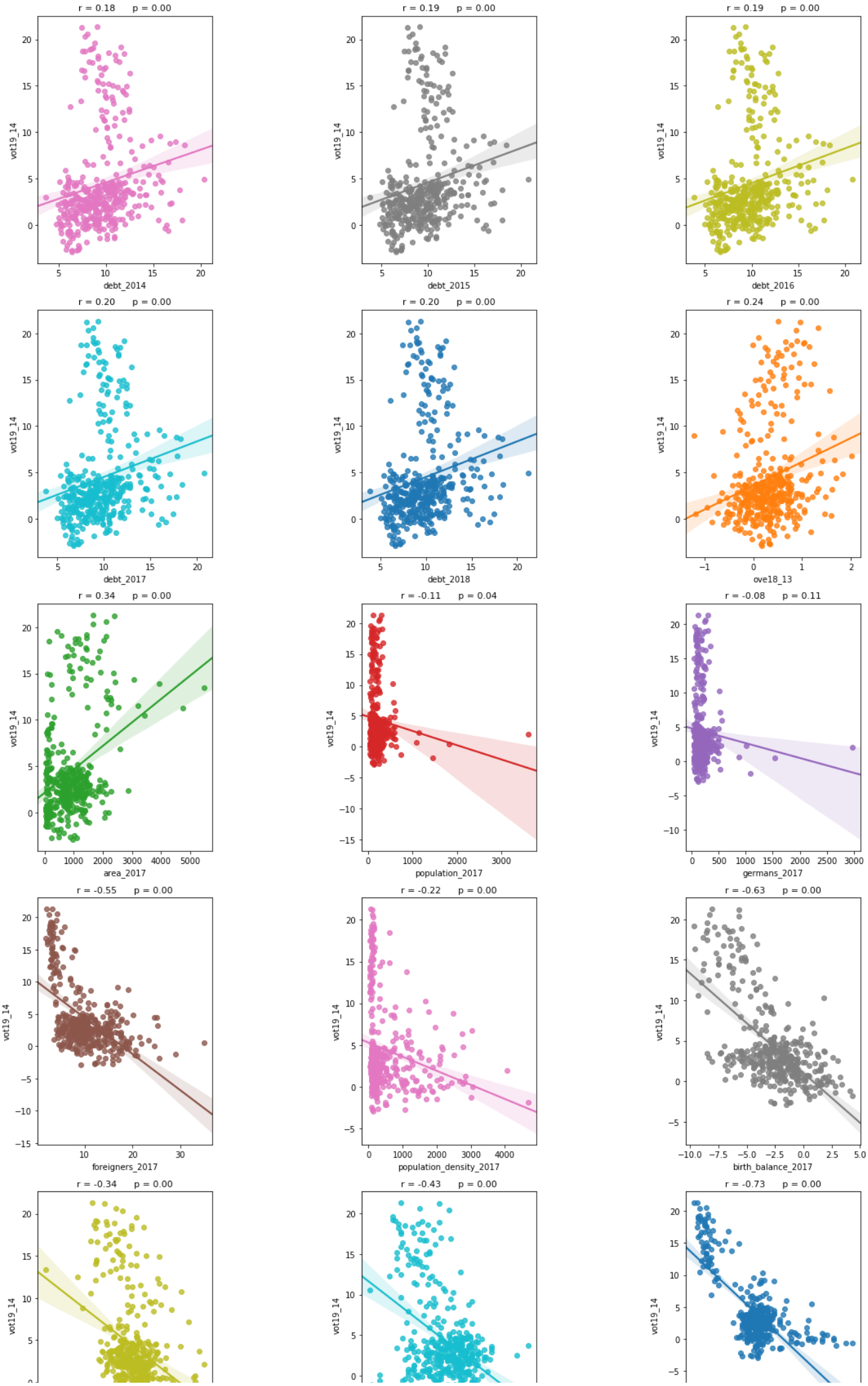
```
numerical_feats = df1.dtypes[df1.dtypes != "object"].index
print("Number of Numerical features: ", len(numerical_feats))

categorical_feats = df1.dtypes[df1.dtypes == "object"].index
print("Number of Categorical features: ", len(categorical_feats))
```

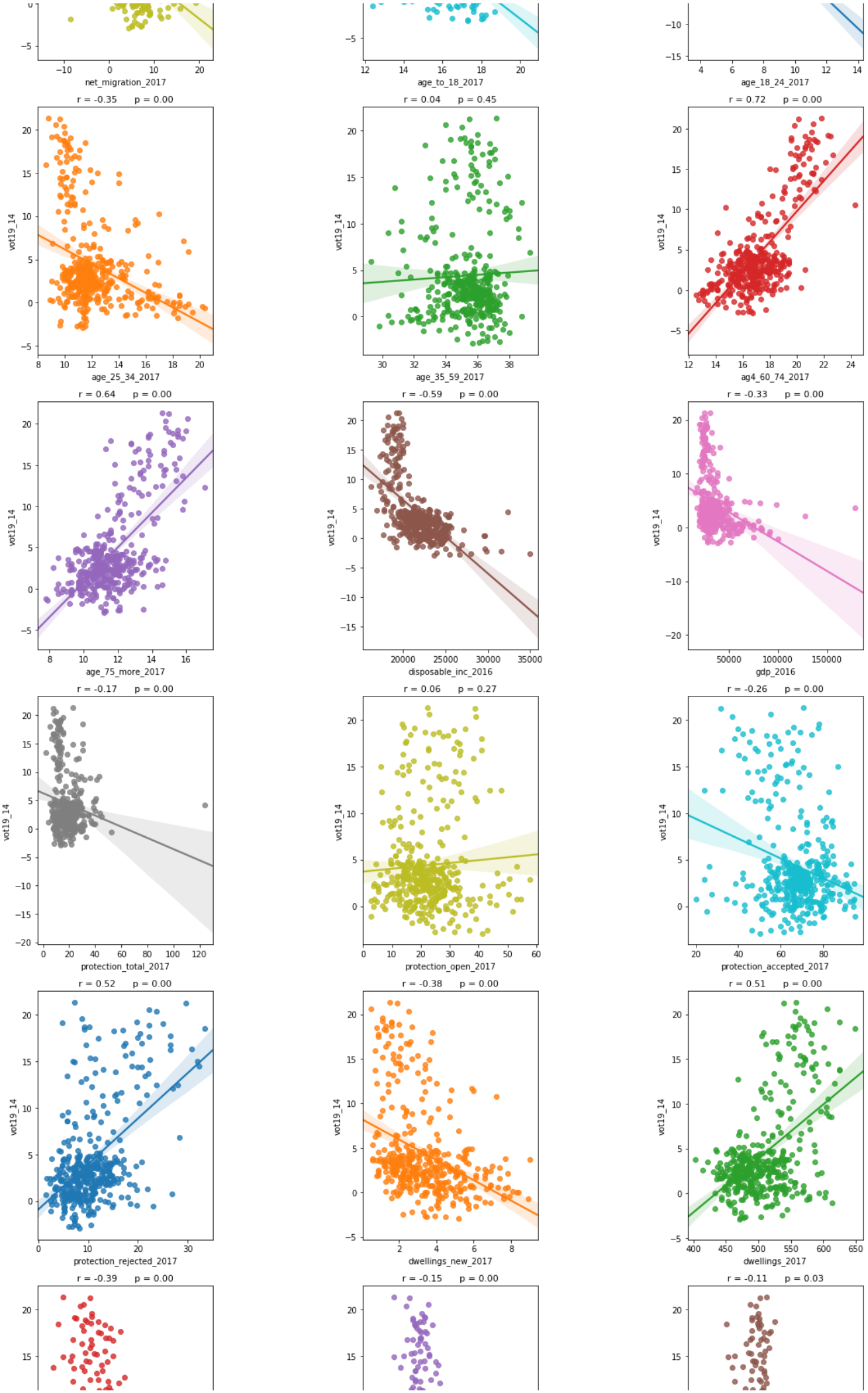
```
Number of Numerical features:  148
Number of Categorical features:  3
```

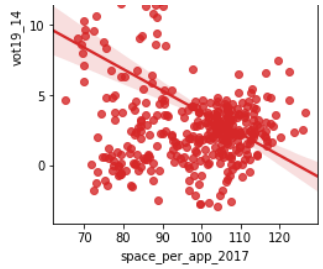
In [152]:

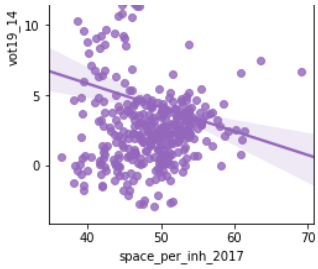
```
target = 'vot19_14'
nr_rows = 55
nr_cols = 3

fig, axs = plt.subplots(nr_rows, nr_cols, figsize=(nr_cols*5,nr_rows*5))

li_num_feats = list(numerical_feats)
li_not_plot = ['state', 'vote19_14', 'turnout19']
li_plot_num_feats = [c for c in list(numerical_feats) if c not in li_not_plot]


for r in range(0,nr_rows):
    for c in range(0,nr_cols):
        i = r*nr_cols+c
        if i < len(li_plot_num_feats):
            sns.regplot(df1[li_plot_num_feats[i]], df1[target], ax = axs[r][c])
            stp = stats.pearsonr(df1[li_plot_num_feats[i]], df1[target])
            #axs[r][c].text(0.4,0.9,"title",fontsize=7)
            str_title = "r = " + "{0:.2f}".format(stp[0]) + "      " "p = " + "{0:.2f}".format(stp[1
])
            axs[r][c].set_title(str_title,fontsize=11)

plt.tight_layout()
plt.show()
```

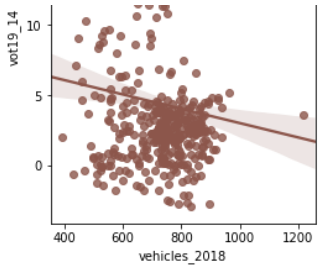


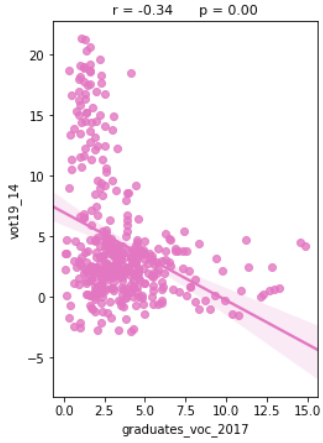

r = 0.18 p = 0.00
vot19_14
debt_2014
r = 0.19 p = 0.00
vot19_14
debt_2015
r = 0.19 p = 0.00
vot19_14
debt_2016
r = 0.20 p = 0.00
vot19_14
debt_2017
r = 0.20 p = 0.00
vot19_14
debt_2018
r = 0.24 p = 0.00
vot19_14
ove18_13
r = 0.34 p = 0.00
vot19_14
area_2017
r = -0.11 p = 0.04
vot19_14
population_2017
r = -0.08 p = 0.11
vot19_14
germans_2017
r = -0.55 p = 0.00
vot19_14
foreigners_2017
r = -0.22 p = 0.00
vot19_14
population_density_2017
r = -0.63 p = 0.00
vot19_14
birth_balance_2017
r = -0.34 p = 0.00
vot19_14
r = -0.43 p = 0.00
vot19_14
r = -0.73 p = 0.00
vot19_14

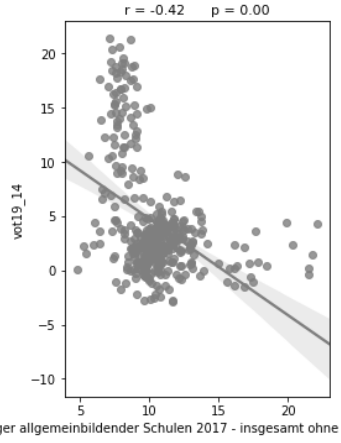

net_migration_2017
age_to_18_2017
age_18_24_2017
r = -0.35 p = 0.00
r = 0.04 p = 0.45
r = 0.72 p = 0.00
vot19_14
age_25_34_2017
age_35_59_2017
ag4_60_74_2017
r = 0.64 p = 0.00
r = -0.59 p = 0.00
r = -0.33 p = 0.00
age_75_more_2017
disposable_inc_2016
gdp_2016
r = -0.17 p = 0.00
r = 0.06 p = 0.27
r = -0.26 p = 0.00
protection_total_2017
protection_open_2017
protection_accepted_2017
r = 0.52 p = 0.00
r = -0.38 p = 0.00
r = 0.51 p = 0.00
protection_rejected_2017
dwellings_new_2017
dwellings_2017
r = -0.39 p = 0.00
r = -0.15 p = 0.00
r = -0.11 p = 0.03

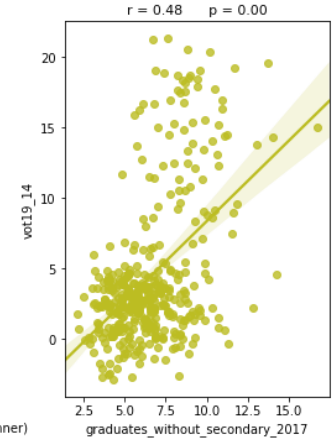




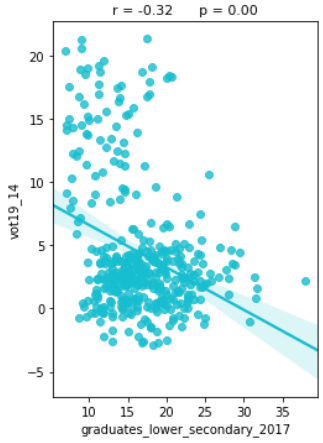


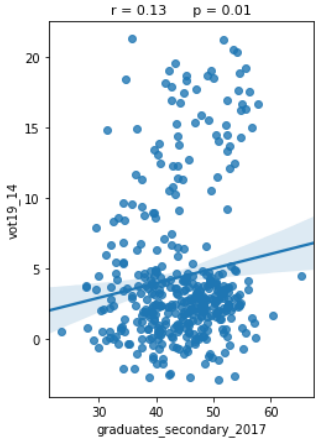


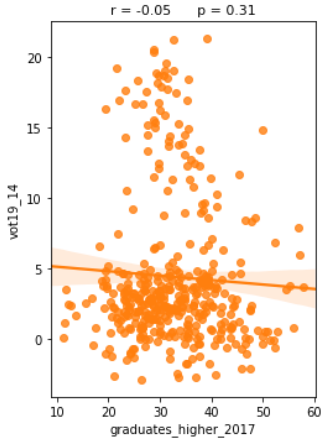


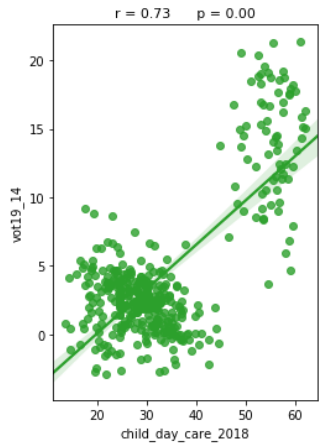


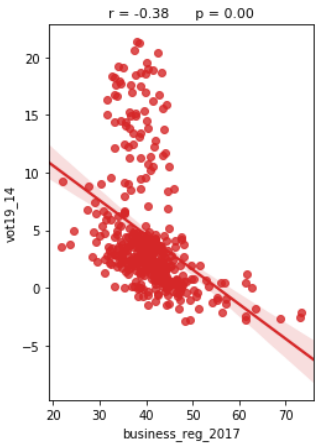
Absolventen/Abgänger allgemeinbildender Schulen 2017 - insgesamt ohne Externe (je 1000 Einwohner)

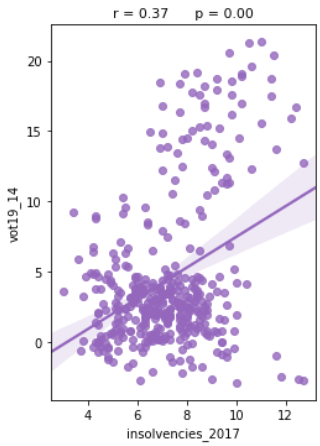


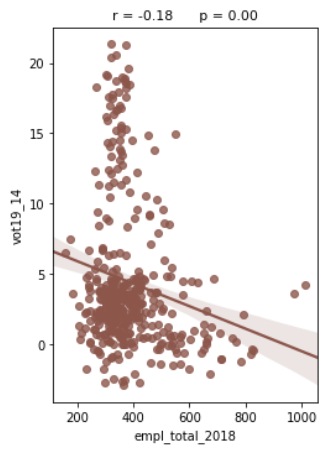


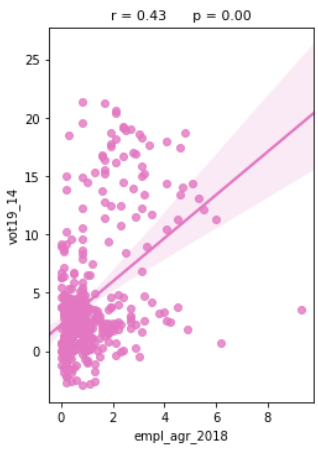


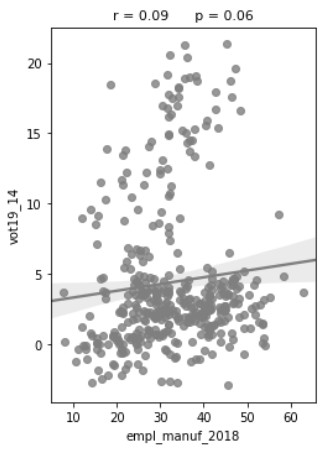


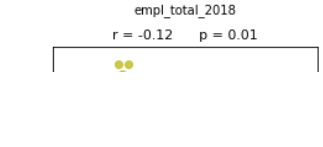


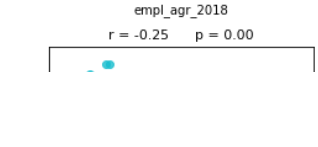


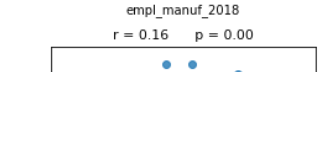


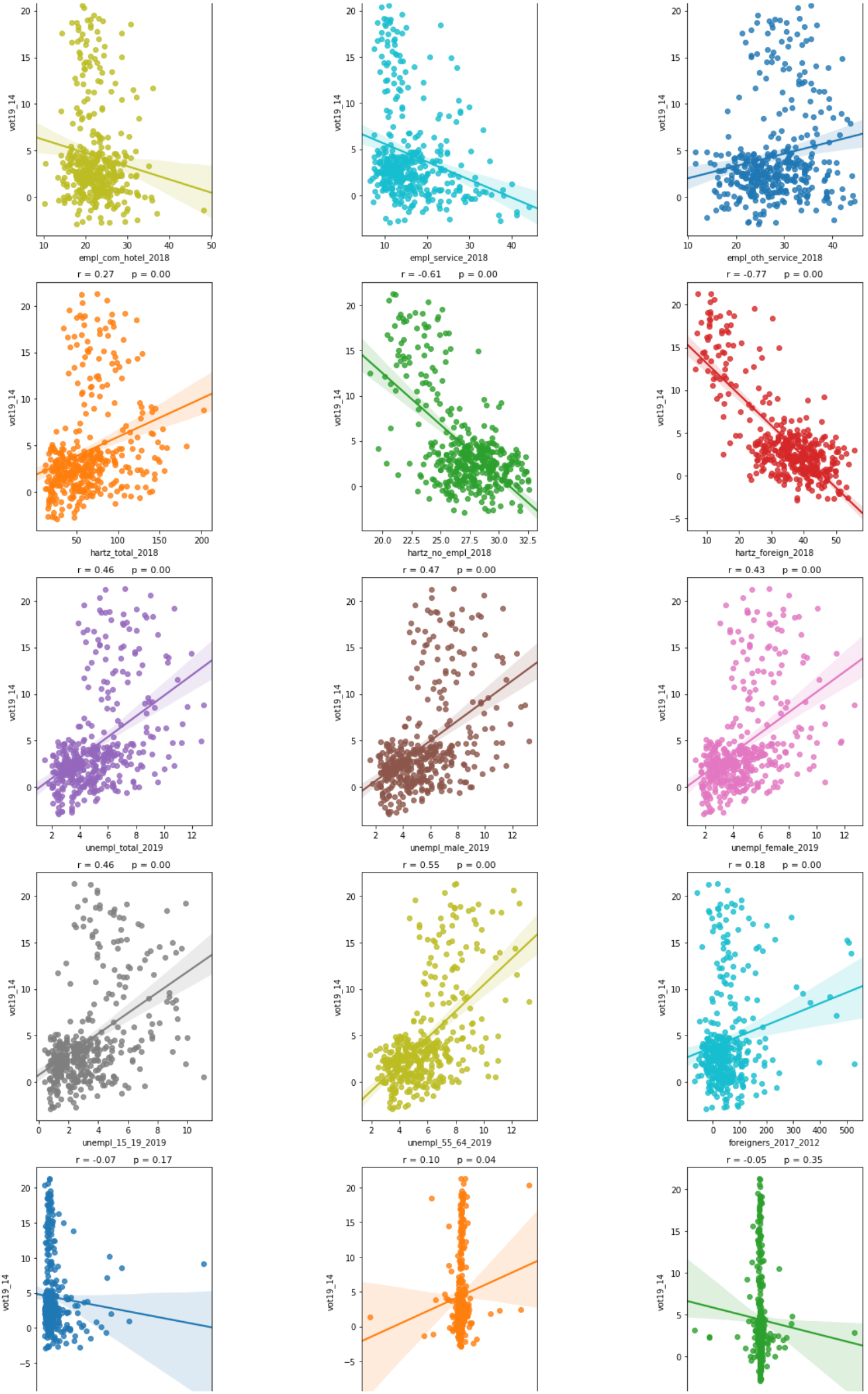


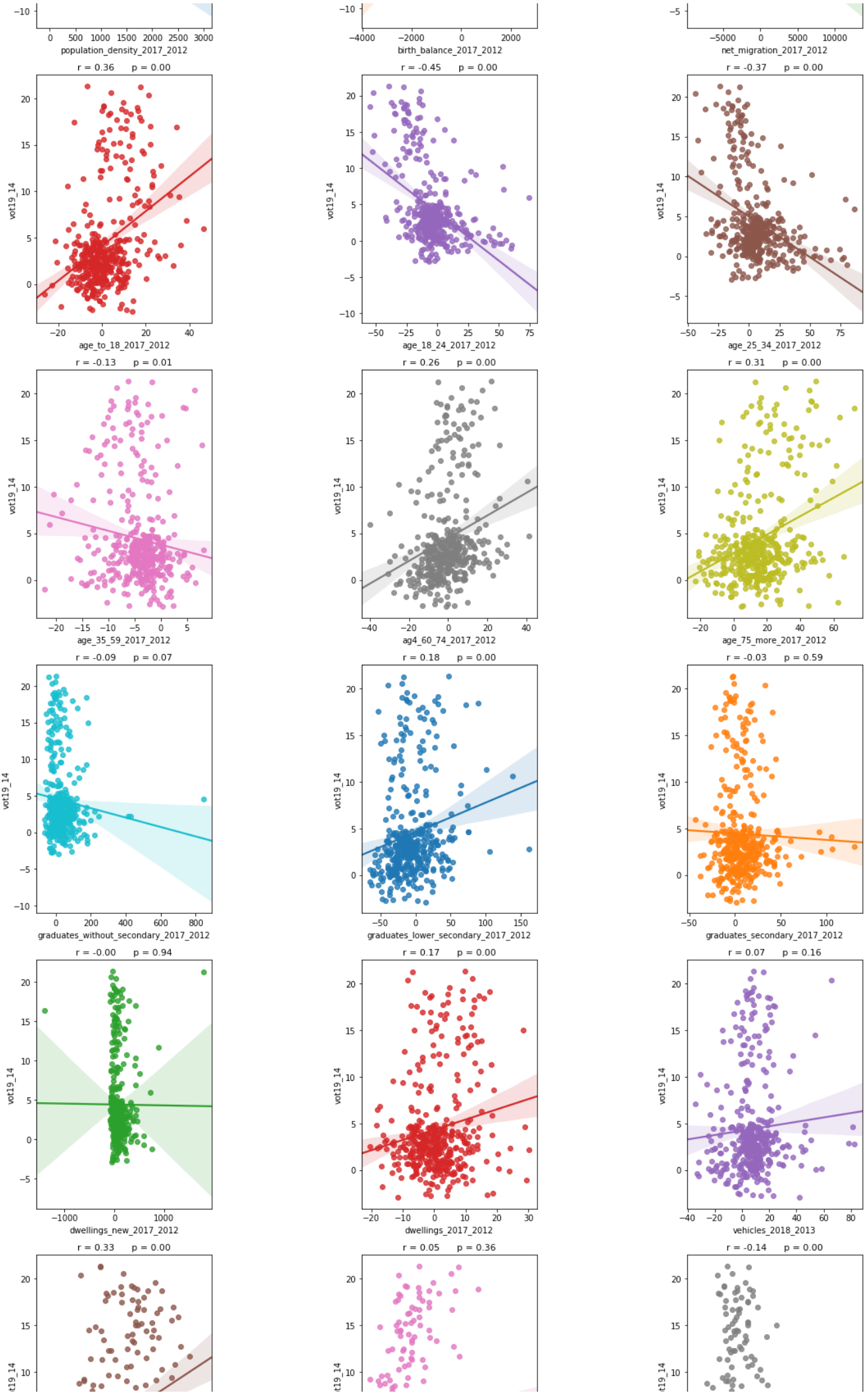


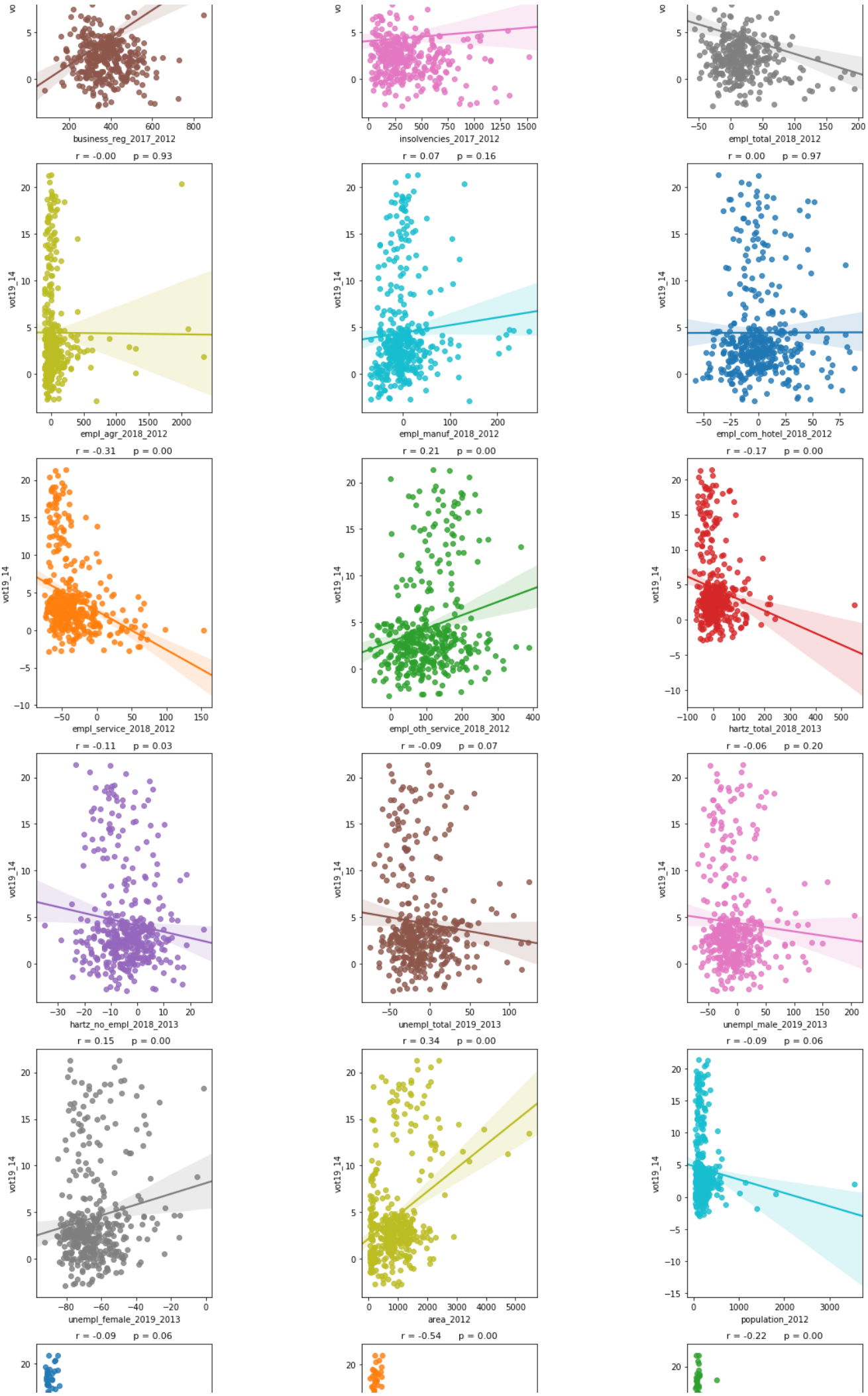


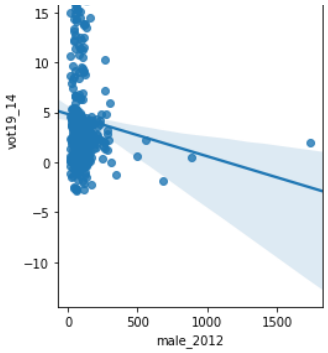


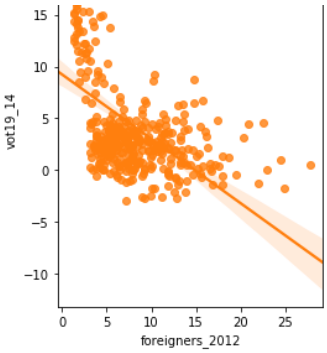


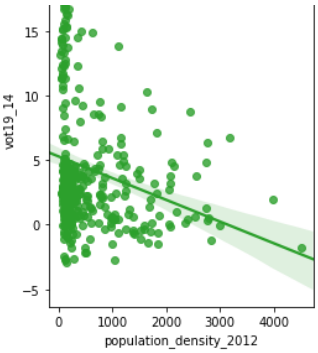

vot19_14
empl_com_hotel_2018
empl_service_2018
empl_oth_service_2018
r = 0.27 p = 0.00
r = -0.61 p = 0.00
r = -0.77 p = 0.00
hartz_total_2018
hartz_no_empl_2018
hartz_foreign_2018
r = 0.46 p = 0.00
r = 0.47 p = 0.00
r = 0.43 p = 0.00
unempl_total_2019
unempl_male_2019
unempl_female_2019
r = 0.46 p = 0.00
r = 0.55 p = 0.00
r = 0.18 p = 0.00
unempl_15_19_2019
unempl_55_64_2019
foreigners_2017_2012
r = -0.07 p = 0.17
r = 0.10 p = 0.04
r = -0.05 p = 0.35

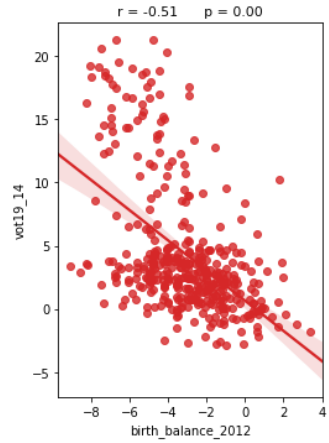

population_density_2017_2012
r = 0.36 p = 0.00
vot19_14
age_to_18_2017_2012
birth_balance_2017_2012
r = -0.45 p = 0.00
age_18_24_2017_2012
net_migration_2017_2012
r = -0.37 p = 0.00
age_25_34_2017_2012
r = -0.13 p = 0.01
age_35_59_2017_2012
r = 0.26 p = 0.00
ag4_60_74_2017_2012
r = 0.31 p = 0.00
age_75_more_2017_2012
r = -0.09 p = 0.07
graduates_without_secondary_2017_2012
r = 0.18 p = 0.00
graduates_lower_secondary_2017_2012
r = -0.03 p = 0.59
graduates_secondary_2017_2012
r = -0.00 p = 0.94
dwellings_new_2017_2012
r = 0.17 p = 0.00
dwellings_2017_2012
r = 0.07 p = 0.16
vehicles_2018_2013
r = 0.33 p = 0.00
r = 0.05 p = 0.36
r = -0.14 p = 0.00

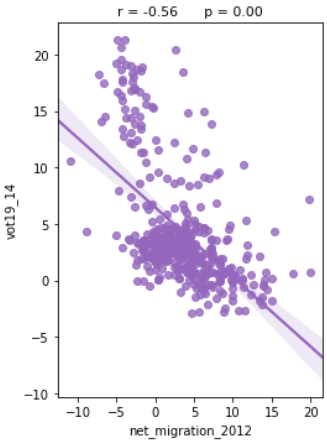

business_reg_2017_2012
r = -0.00 p = 0.93
insolvencies_2017_2012
r = 0.07 p = 0.16
empl_total_2018_2012
r = 0.00 p = 0.97
vot19_14
empl_agr_2018_2012
r = -0.31 p = 0.00
empl_manuf_2018_2012
r = 0.21 p = 0.00
empl_com_hotel_2018_2012
r = -0.17 p = 0.00
empl_service_2018_2012
r = -0.11 p = 0.03
empl_oth_service_2018_2012
r = -0.09 p = 0.07
hartz_total_2018_2013
r = -0.06 p = 0.20
hartz_no_empl_2018_2013
r = 0.15 p = 0.00
unempl_total_2019_2013
r = 0.34 p = 0.00
unempl_male_2019_2013
r = -0.09 p = 0.06
unempl_female_2019_2013
r = -0.09 p = 0.06
area_2012
r = -0.54 p = 0.00
population_2012
r = -0.22 p = 0.00

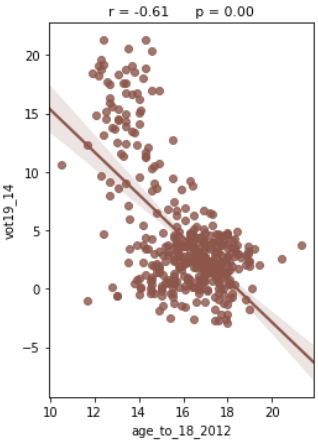

vot19_14
male_2012

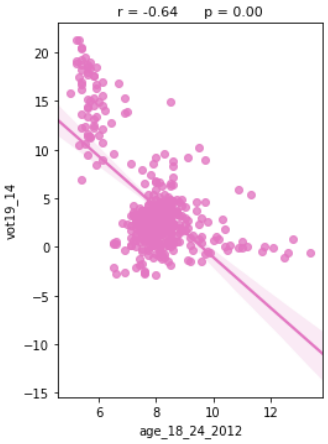

vot19_14
foreigners_2012

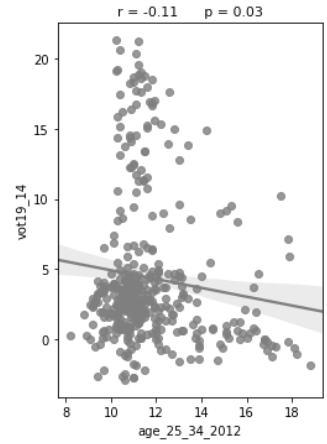

vot19_14
population_density_2012

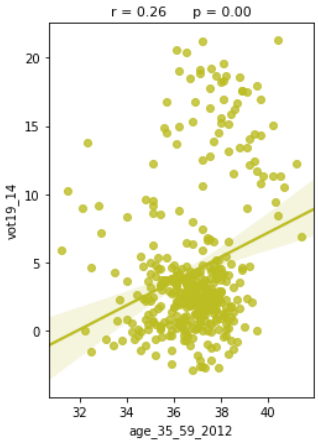

r = -0.51 p = 0.00
vot19_14
birth_balance_2012

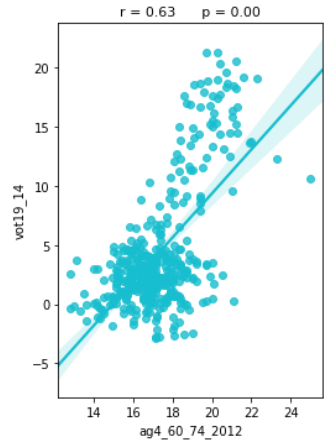

r = -0.56 p = 0.00
vot19_14
net_migration_2012

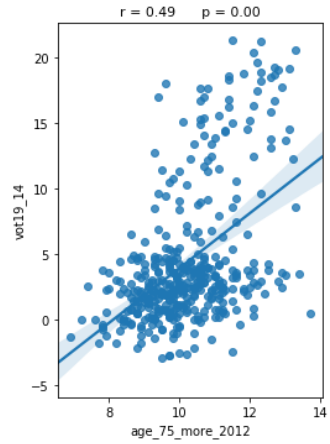

r = -0.61 p = 0.00
vot19_14
age_to_18_2012

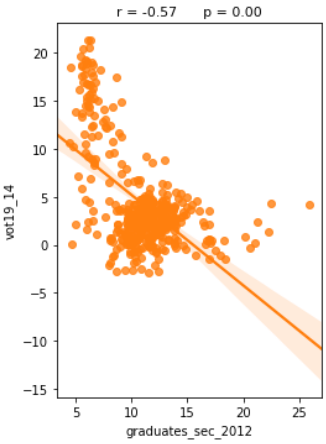

r = -0.64 p = 0.00
vot19_14
age_18_24_2012

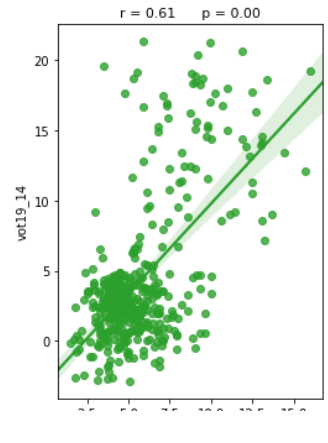

r = -0.11 p = 0.03
vot19_14
age_25_34_2012

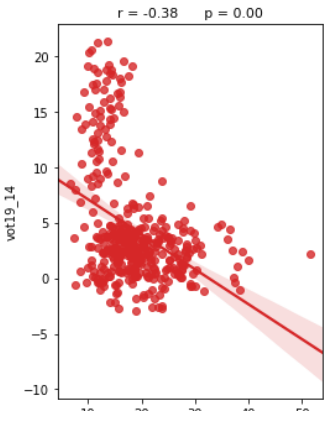

r = 0.26 p = 0.00
vot19_14
age_35_59_2012

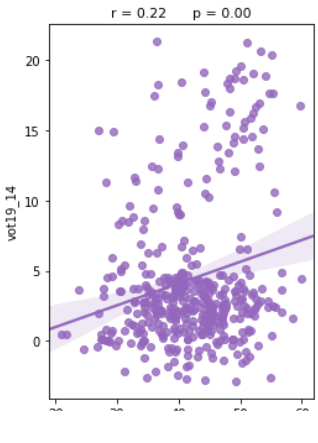

r = 0.63 p = 0.00
vot19_14
ag4_60_74_2012

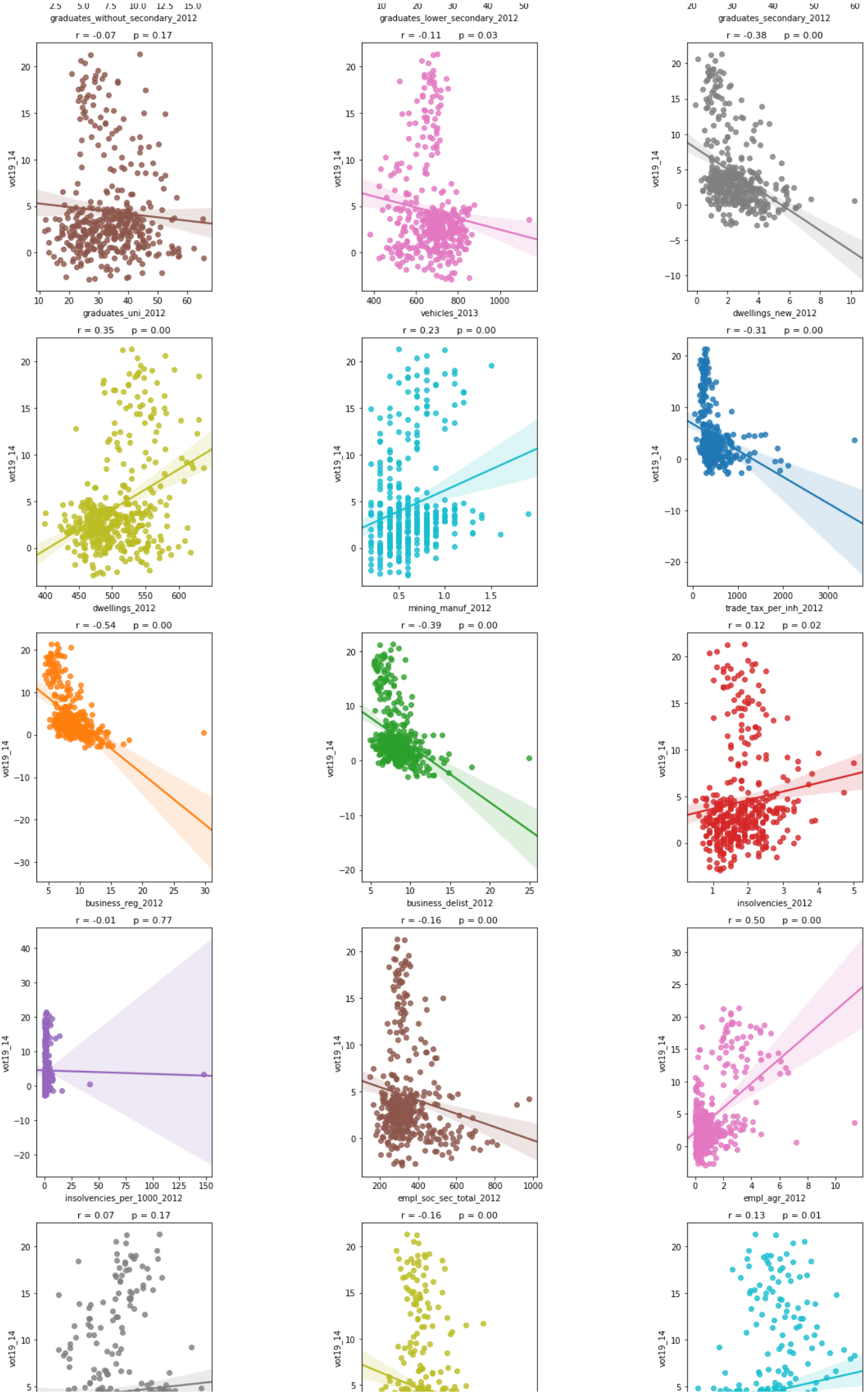

r = 0.49 p = 0.00
vot19_14
age_75_more_2012

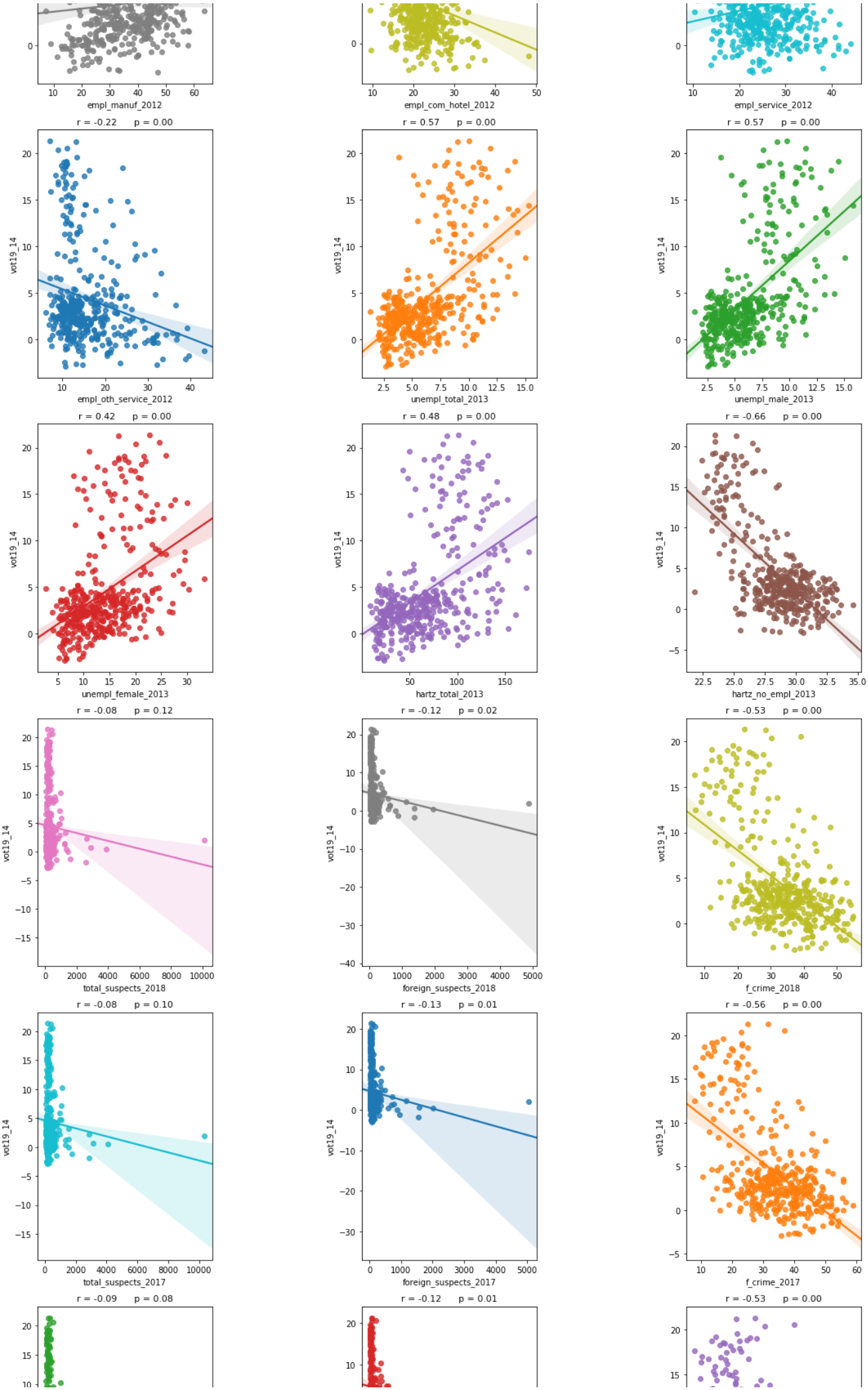

r = -0.57 p = 0.00
vot19_14
graduates_sec_2012

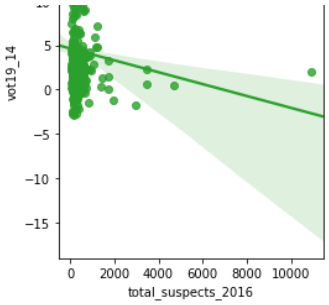

r = 0.61 p = 0.00
vot19_14

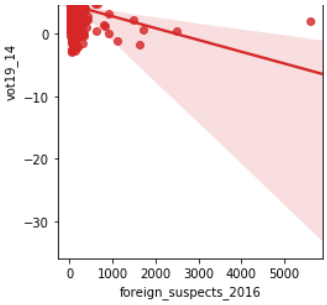

r = -0.38 p = 0.00
vot19_14

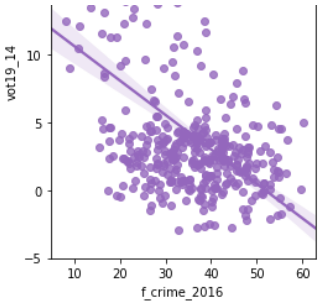

r = 0.22 p = 0.00
vot19_14

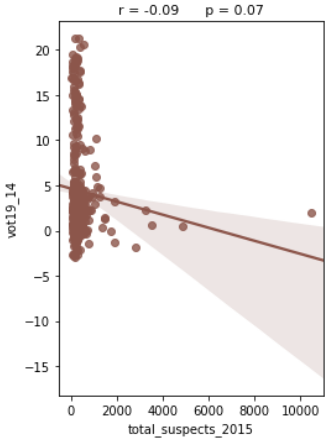

graduates_without_secondary_2012
r = -0.07 p = 0.17
graduates_lower_secondary_2012
r = -0.11 p = 0.03
graduates_secondary_2012
r = -0.38 p = 0.00
graduates_uni_2012
r = 0.35 p = 0.00
vehicles_2013
r = 0.23 p = 0.00
dwellings_new_2012
r = -0.31 p = 0.00
dwellings_2012
r = -0.54 p = 0.00
mining_manuf_2012
r = -0.39 p = 0.00
trade_tax_per_inh_2012
r = 0.12 p = 0.02
business_reg_2012
r = -0.01 p = 0.77
business_delist_2012
r = -0.16 p = 0.00
insolvencies_2012
r = 0.50 p = 0.00
insolvencies_per_1000_2012
r = 0.07 p = 0.17
empl_soc_sec_total_2012
r = -0.16 p = 0.00
empl_agr_2012
r = 0.13 p = 0.01
vot19_14

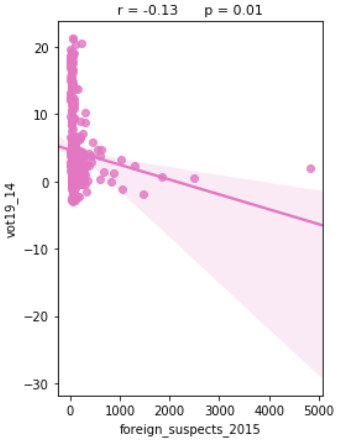

empl_manuf_2012
empl_com_hotel_2012
empl_service_2012
r = -0.22 p = 0.00
r = 0.57 p = 0.00
r = 0.57 p = 0.00
vot19_14
empl_oth_service_2012
unempl_total_2013
unempl_male_2013
r = 0.42 p = 0.00
r = 0.48 p = 0.00
r = -0.66 p = 0.00
unempl_female_2013
hartz_total_2013
hartz_no_empl_2013
r = -0.08 p = 0.12
r = -0.12 p = 0.02
r = -0.53 p = 0.00
total_suspects_2018
foreign_suspects_2018
f_crime_2018
r = -0.08 p = 0.10
r = -0.13 p = 0.01
r = -0.56 p = 0.00
total_suspects_2017
foreign_suspects_2017
f_crime_2017
r = -0.09 p = 0.08
r = -0.12 p = 0.01
r = -0.53 p = 0.00

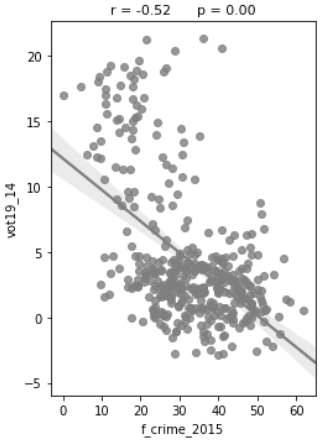

vot19_14
total_suspects_2016

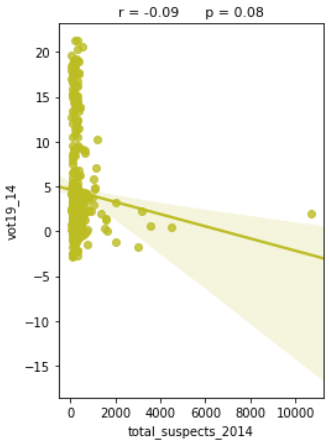

vot19_14
foreign_suspects_2016

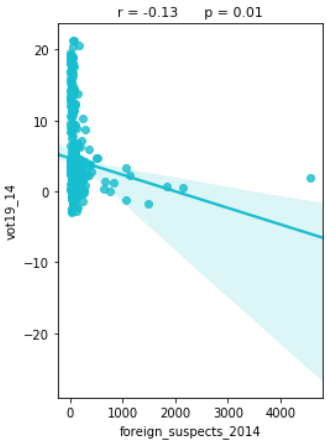

vot19_14
f_crime_2016

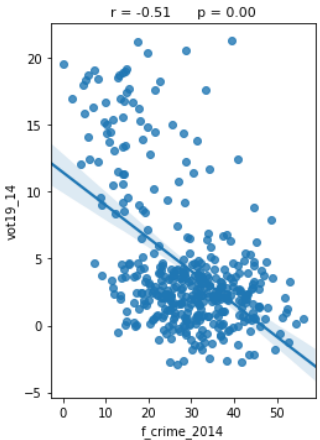

r = -0.09 p = 0.07
vot19_14
total_suspects_2015

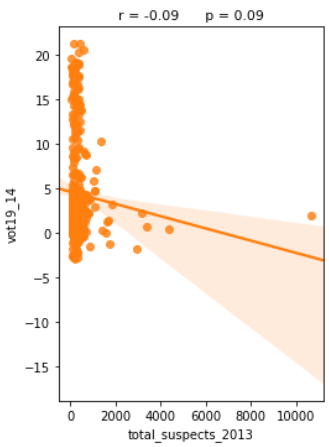

r = -0.13 p = 0.01
vot19_14
foreign_suspects_2015

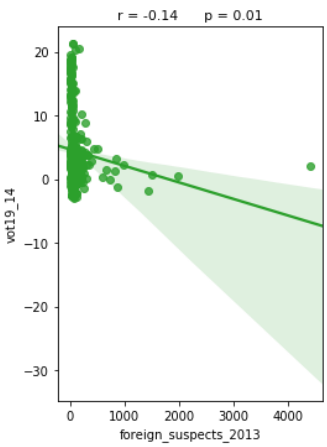

r = -0.52 p = 0.00
vot19_14
f_crime_2015

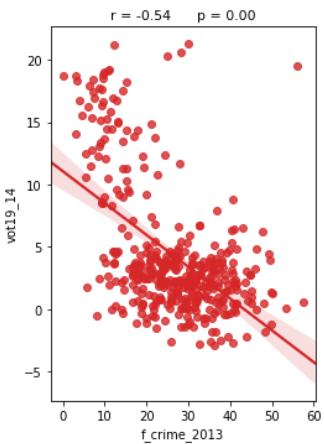

r = -0.09 p = 0.08
vot19_14
total_suspects_2014

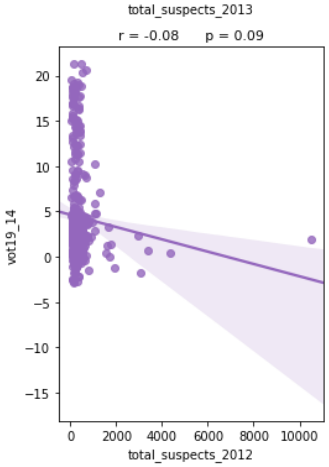

r = -0.13 p = 0.01
vot19_14
foreign_suspects_2014

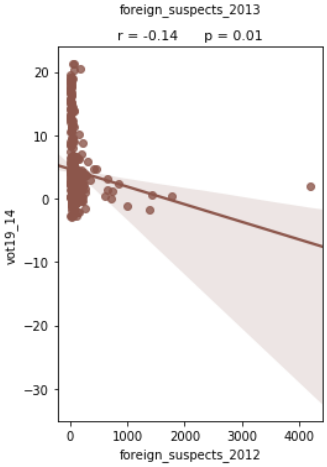

r = -0.51 p = 0.00
vot19_14
f_crime_2014

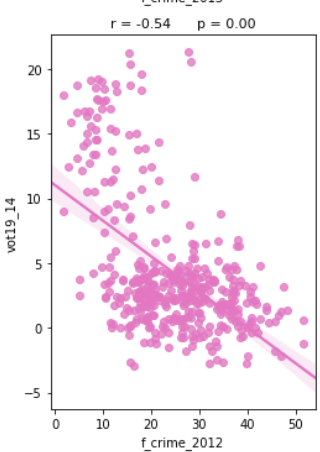

r = -0.09 p = 0.09
vot19_14
total_suspects_2013

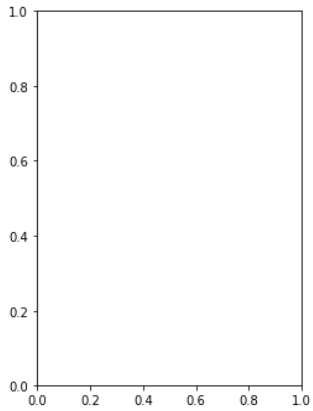
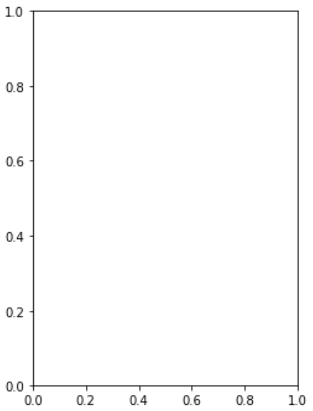
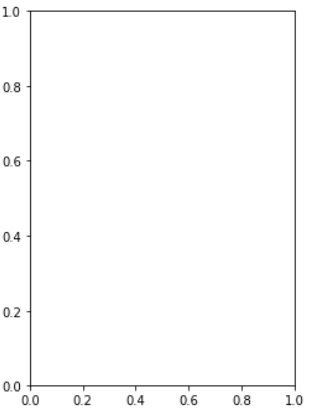
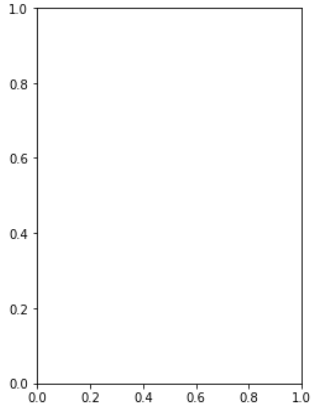
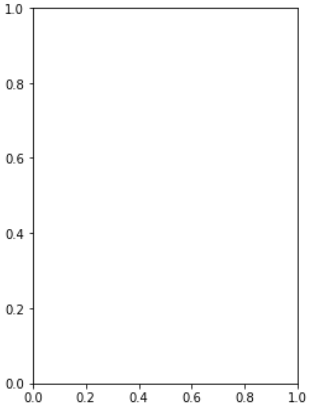
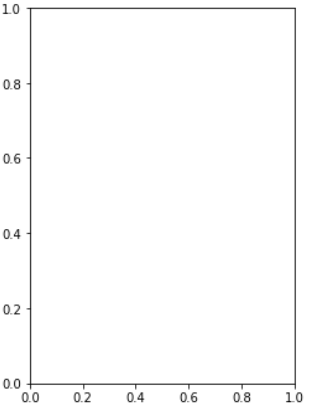
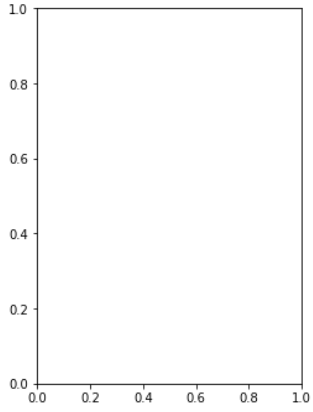
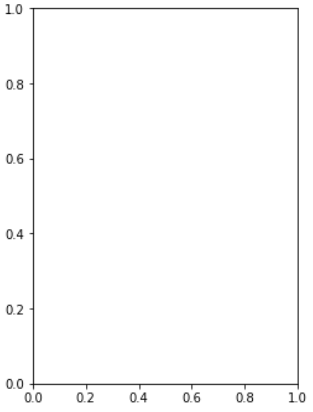
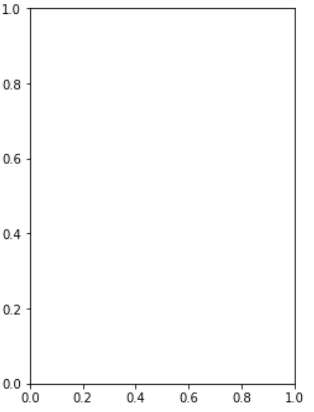
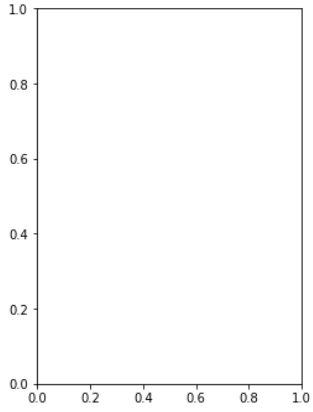
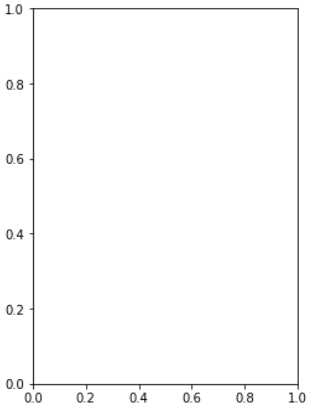
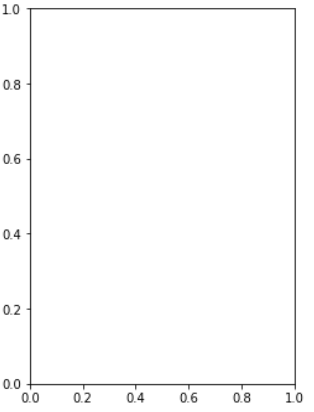
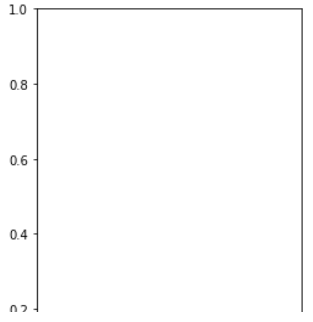
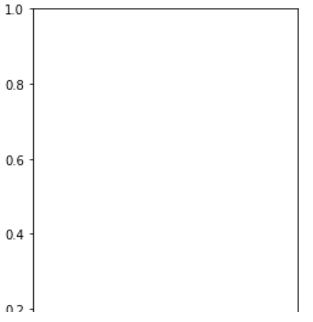
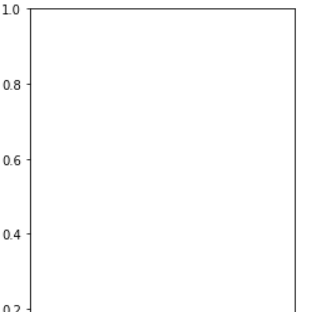

r = -0.14 p = 0.01
vot19_14
foreign_suspects_2013

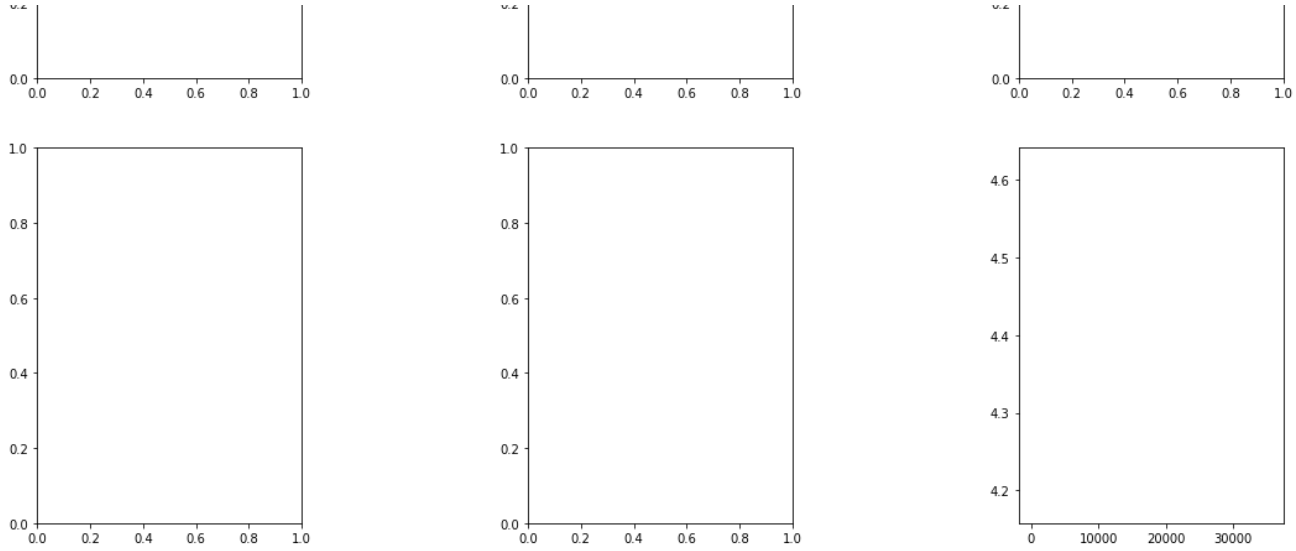

r = -0.54 p = 0.00
vot19_14
f_crime_2013


r = -0.08 p = 0.09
vot19_14
total_suspects_2012


r = -0.14 p = 0.01
vot19_14
foreign_suspects_2012


r = -0.54 p = 0.00
vot19_14
f_crime_2012



- We see that for some features like 'age4_60_74_2017' there is a strong linear correlation (0.72) to the target.
- the features like 'age35_59_2017' the correlation is very weak.
- For this kernel I decided to use only those features for prediction that have a correlation larger than a threshold value to dependent variable.
- This threshold value can be choosen in the coming steps.

In [6]:

```
target = 'vot19_14'
for col in numerical_feats:
    print('{:15}'.format(col),
          'Skewness: {:05.2f}'.format(df1[col].skew()) ,
          '   ' ,
          'Kurtosis: {:06.2f}'.format(df1[col].kurt())
         )
```

```
Nr              Skewness: 00.34     Kurtosis: -00.28
subregion       Skewness: 00.42     Kurtosis: -00.28
vot19_14        Skewness: 01.51     Kurtosis: 001.47
turnout14       Skewness: -0.02     Kurtosis: -00.28
turnout19       Skewness: -0.09     Kurtosis: -00.16
turnout19_14    Skewness: 00.18     Kurtosis: -00.96
debt_2013       Skewness: 00.79     Kurtosis: 000.94
debt_2014       Skewness: 00.80     Kurtosis: 000.95
debt_2015       Skewness: 00.85     Kurtosis: 001.09
debt_2016       Skewness: 00.84     Kurtosis: 001.02
debt_2017       Skewness: 00.89     Kurtosis: 001.10
debt_2018       Skewness: 00.91     Kurtosis: 001.20
ove18_13        Skewness: 00.27     Kurtosis: 000.76
area_2017       Skewness: 01.71     Kurtosis: 006.23
population_2017 Skewness: 08.51     Kurtosis: 104.70
germans_2017    Skewness: 08.34     Kurtosis: 102.57
foreigners_2017 Skewness: 00.90     Kurtosis: 001.32
population_density_2017 Skewness: 02.31     Kurtosis: 006.15
birth_balance_2017 Skewness: -0.21     Kurtosis: -00.14
net_migration_2017 Skewness: 00.20     Kurtosis: 002.20
age_to_18_2017  Skewness: -0.27     Kurtosis: 000.07
age_18_24_2017  Skewness: 00.18     Kurtosis: 001.58
age_25_34_2017  Skewness: 01.47     Kurtosis: 002.00
age_35_59_2017  Skewness: -0.90     Kurtosis: 001.10
ag4_60_74_2017  Skewness: 00.38     Kurtosis: 000.20
age_75_more_2017 Skewness: 00.56     Kurtosis: 000.27
disposable_inc_2016 Skewness: 00.97     Kurtosis: 002.91
gdp_2016        Skewness: 03.48     Kurtosis: 020.46
protection_total_2017 Skewness: 04.01     Kurtosis: 039.29
protection_open_2017 Skewness: 00.75     Kurtosis: 001.01
protection_accepted_2017 Skewness: -0.63     Kurtosis: 000.95
protection_rejected_2017 Skewness: 01.29     Kurtosis: 002.11
dwellings_new_2017 Skewness: 00.78     Kurtosis: 000.17
dwellings_2017  Skewness: 00.52     Kurtosis: -00.23
space_per_app_2017 Skewness: -0.16     Kurtosis: -01.02
space_per_inh_2017 Skewness: 00.33     Kurtosis: 000.62
```

vehicles_2018 Skewness: -0.29 Kurtosis: 000.68
graduates_voc_2017 Skewness: 01.78 Kurtosis: 004.18
Absolventen/Abgänger allgemeinbildender Schulen 2017 - insgesamt ohne Externe (je 1000 Einwohner)
Skewness: 01.58 Kurtosis: 004.79
graduates_without_secondary_2017 Skewness: 00.80 Kurtosis: 001.00
graduates_lower_secondary_2017 Skewness: 00.60 Kurtosis: 000.61
graduates_secondary_2017 Skewness: -0.23 Kurtosis: -00.35
graduates_higher_2017 Skewness: 00.33 Kurtosis: 000.12
child_day_care_2018 Skewness: 00.93 Kurtosis: -00.13
business_reg_2017 Skewness: 01.15 Kurtosis: 004.12
insolvencies_2017 Skewness: 00.37 Kurtosis: 000.10
empl_total_2018 Skewness: 01.76 Kurtosis: 004.30
empl_agr_2018 Skewness: 02.14 Kurtosis: 006.60
empl_manuf_2018 Skewness: 00.10 Kurtosis: -00.51
empl_com_hotel_2018 Skewness: 00.88 Kurtosis: 002.46
empl_service_2018 Skewness: 01.44 Kurtosis: 002.07
empl_oth_service_2018 Skewness: 00.18 Kurtosis: -00.37
hartz_total_2018 Skewness: 00.95 Kurtosis: 000.68
hartz_no_empl_2018 Skewness: -0.43 Kurtosis: -00.24
hartz_foreign_2018 Skewness: -0.65 Kurtosis: -00.31
unempl_total_2019 Skewness: 00.91 Kurtosis: 000.53
unempl_male_2019 Skewness: 00.84 Kurtosis: 000.36
unempl_female_2019 Skewness: 00.96 Kurtosis: 000.78
unempl_15_19_2019 Skewness: 01.13 Kurtosis: 000.87
unempl_55_64_2019 Skewness: 00.87 Kurtosis: 000.62
foreigners_2017_2012 Skewness: 02.82 Kurtosis: 011.26
population_density_2017_2012 Skewness: 05.62 Kurtosis: 046.70
birth_balance_2017_2012 Skewness: -1.14 Kurtosis: 059.14
net_migration_2017_2012 Skewness: 01.74 Kurtosis: 062.35
age_to_18_2017_2012 Skewness: 00.89 Kurtosis: 001.55
age_18_24_2017_2012 Skewness: 00.91 Kurtosis: 002.82
age_25_34_2017_2012 Skewness: 01.06 Kurtosis: 002.79
age_35_59_2017_2012 Skewness: -0.72 Kurtosis: 001.45
ag4_60_74_2017_2012 Skewness: 00.20 Kurtosis: 001.21
age_75_more_2017_2012 Skewness: 00.48 Kurtosis: 000.52
graduates_without_secondary_2017_2012 Skewness: 04.62 Kurtosis: 041.48
graduates_lower_secondary_2017_2012 Skewness: 01.24 Kurtosis: 003.54
graduates_secondary_2017_2012 Skewness: 01.52 Kurtosis: 005.80
dwellings_new_2017_2012 Skewness: 01.94 Kurtosis: 038.91
dwellings_2017_2012 Skewness: 00.48 Kurtosis: 000.90
vehicles_2018_2013 Skewness: 00.91 Kurtosis: 004.13
business_reg_2017_2012 Skewness: 00.57 Kurtosis: 000.55
insolvencies_2017_2012 Skewness: 01.24 Kurtosis: 001.88
empl_total_2018_2012 Skewness: 01.59 Kurtosis: 004.69
empl_agr_2018_2012 Skewness: 05.86 Kurtosis: 042.70
empl_manuf_2018_2012 Skewness: 02.46 Kurtosis: 009.42
empl_com_hotel_2018_2012 Skewness: 00.78 Kurtosis: 001.05
empl_service_2018_2012 Skewness: 01.88 Kurtosis: 004.87
empl_oth_service_2018_2012 Skewness: 00.49 Kurtosis: 000.31
hartz_total_2018_2013 Skewness: 03.32 Kurtosis: 022.95
hartz_no_empl_2018_2013 Skewness: 00.05 Kurtosis: 000.31
unempl_total_2019_2013 Skewness: 01.26 Kurtosis: 002.59
unempl_male_2019_2013 Skewness: 01.65 Kurtosis: 004.55
unempl_female_2019_2013 Skewness: 01.30 Kurtosis: 002.48
area_2012 Skewness: 01.70 Kurtosis: 006.17
population_2012 Skewness: 08.50 Kurtosis: 104.69
male_2012 Skewness: 08.53 Kurtosis: 105.43
foreigners_2012 Skewness: 01.03 Kurtosis: 001.35
population_density_2012 Skewness: 02.30 Kurtosis: 006.01
birth_balance_2012 Skewness: -0.02 Kurtosis: -00.31
net_migration_2012 Skewness: 00.34 Kurtosis: 000.34
age_to_18_2012 Skewness: -0.42 Kurtosis: -00.33
age_18_24_2012 Skewness: 00.42 Kurtosis: 001.94
age_25_34_2012 Skewness: 01.55 Kurtosis: 002.25
age_35_59_2012 Skewness: -0.45 Kurtosis: 001.25
ag4_60_74_2012 Skewness: 00.48 Kurtosis: 000.75
age_75_more_2012 Skewness: 00.30 Kurtosis: -00.11
graduates_sec_2012 Skewness: 00.62 Kurtosis: 001.74
graduates_without_secondary_2012 Skewness: 01.29 Kurtosis: 001.84
graduates_lower_secondary_2012 Skewness: 01.08 Kurtosis: 001.90
graduates_secondary_2012 Skewness: -0.13 Kurtosis: -00.43
graduates_uni_2012 Skewness: 00.35 Kurtosis: -00.08
vehicles_2013 Skewness: -0.22 Kurtosis: 000.78
dwellings_new_2012 Skewness: 01.16 Kurtosis: 002.30
dwellings_2012 Skewness: 00.47 Kurtosis: -00.05
mining_manuf_2012 Skewness: 00.95 Kurtosis: 001.36
trade_tax_per_inh_2012 Skewness: 04.10 Kurtosis: 027.50

```
business_reg_2012 Skewness: 02.31     Kurtosis: 015.33
business_delist_2012 Skewness: 02.18     Kurtosis: 012.15
insolvencies_2012 Skewness: 01.00     Kurtosis: 001.73
insolvencies_per_1000_2012 Skewness: 17.05     Kurtosis: 315.28
empl_soc_sec_total_2012 Skewness: 01.86     Kurtosis: 004.85
empl_agr_2012    Skewness: 02.43     Kurtosis: 008.97
empl_manuf_2012 Skewness: 00.06     Kurtosis: -00.49
empl_com_hotel_2012 Skewness: 00.80     Kurtosis: 002.28
empl_service_2012 Skewness: 00.34     Kurtosis: -00.04
empl_oth_service_2012 Skewness: 01.44     Kurtosis: 001.97
unempl_total_2013 Skewness: 00.78     Kurtosis: -00.05
unempl_male_2013 Skewness: 00.81     Kurtosis: 000.09
unempl_female_2013 Skewness: 00.62     Kurtosis: -00.11
hartz_total_2013 Skewness: 00.62     Kurtosis: -00.38
hartz_no_empl_2013 Skewness: -0.35     Kurtosis: -00.36
total_suspects_2018 Skewness: 11.81     Kurtosis: 177.78
foreign_suspects_2018 Skewness: 11.93     Kurtosis: 178.61
f_crime_2018    Skewness: -0.14     Kurtosis: -00.39
total_suspects_2017 Skewness: 11.38     Kurtosis: 166.68
foreign_suspects_2017 Skewness: 11.48     Kurtosis: 167.13
f_crime_2017    Skewness: -0.17     Kurtosis: -00.52
total_suspects_2016 Skewness: 10.81     Kurtosis: 151.68
foreign_suspects_2016 Skewness: 11.01     Kurtosis: 155.13
f_crime_2016    Skewness: -0.12     Kurtosis: -00.61
total_suspects_2015 Skewness: 10.48     Kurtosis: 143.24
foreign_suspects_2015 Skewness: 09.92     Kurtosis: 127.19
f_crime_2015    Skewness: -0.09     Kurtosis: -00.58
total_suspects_2014 Skewness: 10.64     Kurtosis: 148.31
foreign_suspects_2014 Skewness: 10.04     Kurtosis: 130.21
f_crime_2014    Skewness: -0.08     Kurtosis: -00.50
total_suspects_2013 Skewness: 10.97     Kurtosis: 156.40
foreign_suspects_2013 Skewness: 10.64     Kurtosis: 145.52
f_crime_2013    Skewness: 00.05     Kurtosis: -00.56
total_suspects_2012 Skewness: 10.76     Kurtosis: 151.82
foreign_suspects_2012 Skewness: 10.70     Kurtosis: 147.39
f_crime_2012    Skewness: 00.14     Kurtosis: -00.62
```

In [7]:

```
min_val_corr = 0.5
corr = df1.corr()
corr_abs = corr.abs()

numerical_cols = len(numerical_feats)
high_corr = corr_abs.nlargest(numerical_cols, target)[target]

cols_abv_corr_limit = list(high_corr[high_corr.values > min_val_corr].index)
cols_bel_corr_limit = list(high_corr[high_corr.values <= min_val_corr].index)
```

In [8]:

```
print(high_corr)
print("*"*30)
print("List of numerical features with r above min_val_corr :")
print(cols_abv_corr_limit)
print("*"*30)
print("List of numerical features with r below min_val_corr :")
print(cols_bel_corr_limit)
```

```
vot19_14
000000
hartz_foreign_2018
.768042
age_18_24_2017
.733853
child_day_care_2018
.730496
ag4_60_74_2017
.715892
subregion
0.674840
Nr
665838
hartz_no_empl_2013
.656867
```

.650667
age_18_24_2012
.644635
age_75_more_2017
.643707
ag4_60_74_2012
.634542
birth_balance_2017
.627505
age_to_18_2012
.611187
graduates_without_secondary_2012
0.610416
hartz_no_empl_2018
.608727
disposable_inc_2016
.586712
unempl_male_2013
.574578
graduates_sec_2012
.570598
unempl_total_2013
.566574
net_migration_2012
.558362
f_crime_2017
.556580
unempl_55_64_2019
.554606
foreigners_2017
.545720
foreigners_2012
.544824
f_crime_2012
.539652
business_reg_2012
.538696
f_crime_2013
.535228
f_crime_2016
.532991
f_crime_2018
.527540
f_crime_2015
.524866
protection_rejected_2017
.515118
birth_balance_2012
.513256
f_crime_2014
.511593
dwellings_2017
.507236
empl_agr_2012
.495309
age_75_more_2012
.491049
graduates_without_secondary_2017
0.484883
hartz_total_2013
.478324
unempl_male_2019
.469468
unempl_total_2019
.459352
unempl_15_19_2019
.455095
age_18_24_2017_2012
.446146
unempl_female_2019
.434602
age_to_18_2017
.432441
empl_agr_2018
.428920
Absolventen/Abgänger allgemeinbildender Schulen 2017 - insgesamt ohne Externe (je 1000 Einwohner)
0.420188
unempl_female_2013

unempl_female_2015
.417096
business_delist_2012
.392923
space_per_app_2017
.392173
graduates_lower_secondary_2012
0.383047
dwellings_new_2012
.381514
dwellings_new_2017
.379668
business_reg_2017
.379109
insolvencies_2017
.369648
age_25_34_2017_2012
.369181
turnout19_14
.369137
age_to_18_2017_2012
.364020
dwellings_2012
.353913
age_25_34_2017
.350205
area_2017
0.344429
area_2012
0.343590
graduates_voc_2017
.339728
net_migration_2017
.335366
business_reg_2017_2012
.327242
gdp_2016
327122
graduates_lower_secondary_2017
0.318215
empl_service_2018_2012
.311119
trade_tax_per_inh_2012
.310236
age_75_more_2017_2012
.306074
hartz_total_2018
.268308
protection_accepted_2017
.262290
age_35_59_2012
.255948
ag4_60_74_2017_2012
.255523
empl_service_2018
.247994
ove18_13
236387
mining_manuf_2012
.227839
population_density_2017
.223825
empl_oth_service_2012
.222984
population_density_2012
.218104
graduates_secondary_2012
.215228
empl_oth_service_2018_2012
.205944
debt_2018
0.198412
debt_2017
0.196195
debt_2016
0.192549
debt_2015
0.189063

0.189963
foreigners_2017_2012
.184731
empl_total_2018
.183552
graduates_lower_secondary_2017_2012
0.182641
debt_2014
0.181594
turnout19
0.178154
protection_total_2017
.174435
hartz_total_2018_2013
.171244
dwellings_2017_2012
.170693
empl_com_hotel_2012
.164260
empl_oth_service_2018
.163035
debt_2013
0.159191
empl_soc_sec_total_2012
.156154
space_per_inh_2017
.153039
unempl_female_2019_2013
.148861
empl_total_2018_2012
.143407
foreign_suspects_2012
.139282
foreign_suspects_2013
.137638
graduates_secondary_2017
.134506
empl_service_2012
.133308
foreign_suspects_2015
.133104
foreign_suspects_2014
.131889
age_35_59_2017_2012
.128992
turnout14
0.128768
foreign_suspects_2017
.126743
foreign_suspects_2016
.124757
empl_com_hotel_2018
.122968
insolvencies_2012
.118855
foreign_suspects_2018
.118678
age_25_34_2012
.111508
hartz_no_empl_2018_2013
.110559
vehicles_2013
.110441
vehicles_2018
.105623
population_2017
.105071
birth_balance_2017_2012
.101380
empl_manuf_2018
.094039
population_2012
.093538
male_2012
0.092658
unempl_total_2019_2013
.091080
total_suspects_2015

```
total_suspects_2015
.090458
graduates_without_secondary_2017_2012
0.089109
total_suspects_2014
.086246
total_suspects_2016
.086211
total_suspects_2013
.085550
total_suspects_2012
.083868
total_suspects_2017
.082195
germans_2017
.080996
total_suspects_2018
.078443
empl_manuf_2018_2012
.070412
vehicles_2018_2013
.069607
graduates_uni_2012
.068876
population_density_2017_2012
.068413
empl_manuf_2012
.067874
unempl_male_2019_2013
.064456
protection_open_2017
.055762
graduates_higher_2017
.050933
net_migration_2017_2012
.046752
insolvencies_2017_2012
.045923
age_35_59_2017
.037873
graduates_secondary_2017_2012
.027159
insolvencies_per_1000_2012
.014906
empl_agr_2018_2012
.004121
dwellings_new_2017_2012
.003734
empl_com_hotel_2018_2012
.001781
Name: vot19_14, dtype: float64
******************************
List of numerical features with r above min_val_corr :
['vot19_14', 'hartz_foreign_2018', 'age_18_24_2017', 'child_day_care_2018', 'ag4_60_74_2017', 'sub
region', 'Nr', 'hartz_no_empl_2013', 'age_18_24_2012', 'age_75_more_2017', 'ag4_60_74_2012',
'birth_balance_2017', 'age_to_18_2012', 'graduates_without_secondary_2012', 'hartz_no_empl_2018',
'disposable_inc_2016', 'unempl_male_2013', 'graduates_sec_2012', 'unempl_total_2013',
'net_migration_2012', 'f_crime_2017', 'unempl_55_64_2019', 'foreigners_2017', 'foreigners_2012',
'f_crime_2012', 'business_reg_2012', 'f_crime_2013', 'f_crime_2016', 'f_crime_2018',
'f_crime_2015', 'protection_rejected_2017', 'birth_balance_2012', 'f_crime_2014',
'dwellings_2017']
******************************
List of numerical features with r below min_val_corr :
['empl_agr_2012', 'age_75_more_2012', 'graduates_without_secondary_2017', 'hartz_total_2013', 'une
mpl_male_2019', 'unempl_total_2019', 'unempl_15_19_2019', 'age_18_24_2017_2012',
'unempl_female_2019', 'age_to_18_2017', 'empl_agr_2018', 'Absolventen/Abgänger allgemeinbildender
Schulen 2017 - insgesamt ohne Externe (je 1000 Einwohner)', 'unempl_female_2013',
'business_delist_2012', 'space_per_app_2017', 'graduates_lower_secondary_2012',
'dwellings_new_2012', 'dwellings_new_2017', 'business_reg_2017', 'insolvencies_2017',
'age_25_34_2017_2012', 'turnout19_14', 'age_to_18_2017_2012', 'dwellings_2012', 'age_25_34_2017',
'area_2017', 'area_2012', 'graduates_voc_2017', 'net_migration_2017', 'business_reg_2017_2012', 'g
dp_2016', 'graduates_lower_secondary_2017', 'empl_service_2018_2012', 'trade_tax_per_inh_2012', 'a
ge_75_more_2017_2012', 'hartz_total_2018', 'protection_accepted_2017', 'age_35_59_2012',
'ag4_60_74_2017_2012', 'empl_service_2018', 'ove18_13', 'mining_manuf_2012',
'population_density_2017', 'empl_oth_service_2012', 'population_density_2012',
'graduates_secondary_2012', 'empl_oth_service_2018_2012', 'debt_2018', 'debt_2017', 'debt_2016', '
debt_2015', 'foreigners_2017_2012', 'empl_total_2018', 'graduates_lower_secondary_2017_2012',
```

'debt_2014', 'turnout19', 'protection_total_2017', 'hartz_total_2018_2013', 'dwellings_2017_2012',
'empl_com_hotel_2012', 'empl_oth_service_2018', 'debt_2013', 'empl_soc_sec_total_2012',
'space_per_inh_2017', 'unempl_female_2019_2013', 'empl_total_2018_2012', 'foreign_suspects_2012',
'foreign_suspects_2013', 'graduates_secondary_2017', 'empl_service_2012', 'foreign_suspects_2015',
'foreign_suspects_2014', 'age_35_59_2017_2012', 'turnout14', 'foreign_suspects_2017',
'foreign_suspects_2016', 'empl_com_hotel_2018', 'insolvencies_2012', 'foreign_suspects_2018',
'age_25_34_2012', 'hartz_no_empl_2018_2013', 'vehicles_2013', 'vehicles_2018', 'population_2017',
'birth_balance_2017_2012', 'empl_manuf_2018', 'population_2012', 'male_2012',
'unempl_total_2019_2013', 'total_suspects_2015', 'graduates_without_secondary_2017_2012',
'total_suspects_2014', 'total_suspects_2016', 'total_suspects_2013', 'total_suspects_2012',
'total_suspects_2017', 'germans_2017', 'total_suspects_2018', 'empl_manuf_2018_2012',
'vehicles_2018_2013', 'graduates_uni_2012', 'population_density_2017_2012', 'empl_manuf_2012', 'un
empl_male_2019_2013', 'protection_open_2017', 'graduates_higher_2017', 'net_migration_2017_2012',
'insolvencies_2017_2012', 'age_35_59_2017', 'graduates_secondary_2017_2012',
'insolvencies_per_1000_2012', 'empl_agr_2018_2012', 'dwellings_new_2017_2012',
'empl_com_hotel_2018_2012']

In [100]:

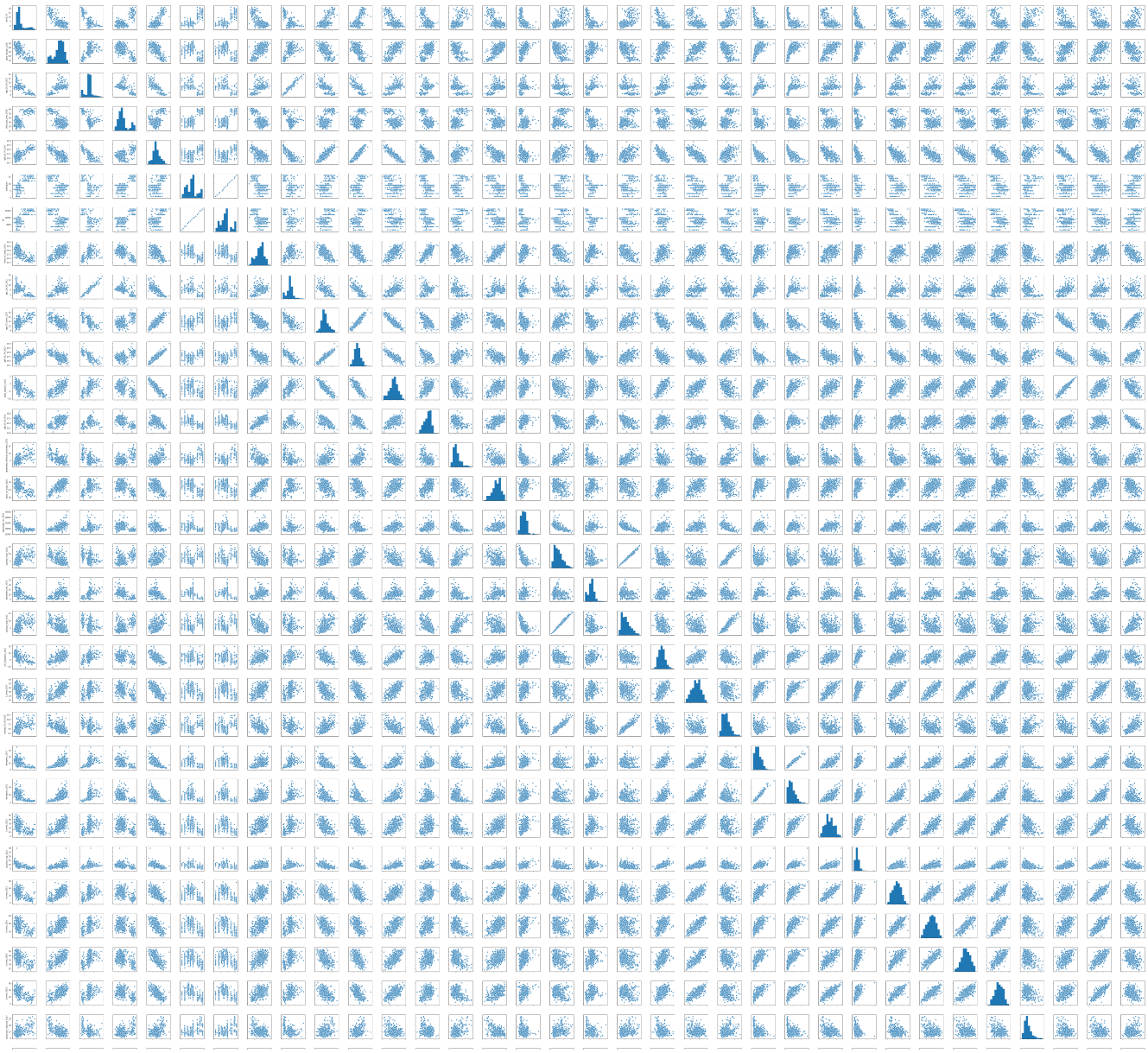
```
df2 = df1[cols_abv_corr_limit]
corr = df2.corr()
```

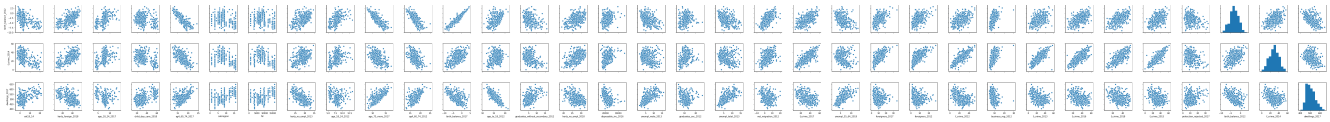
In [15]:

```
sns.pairplot(df2)
```

Out[15]:

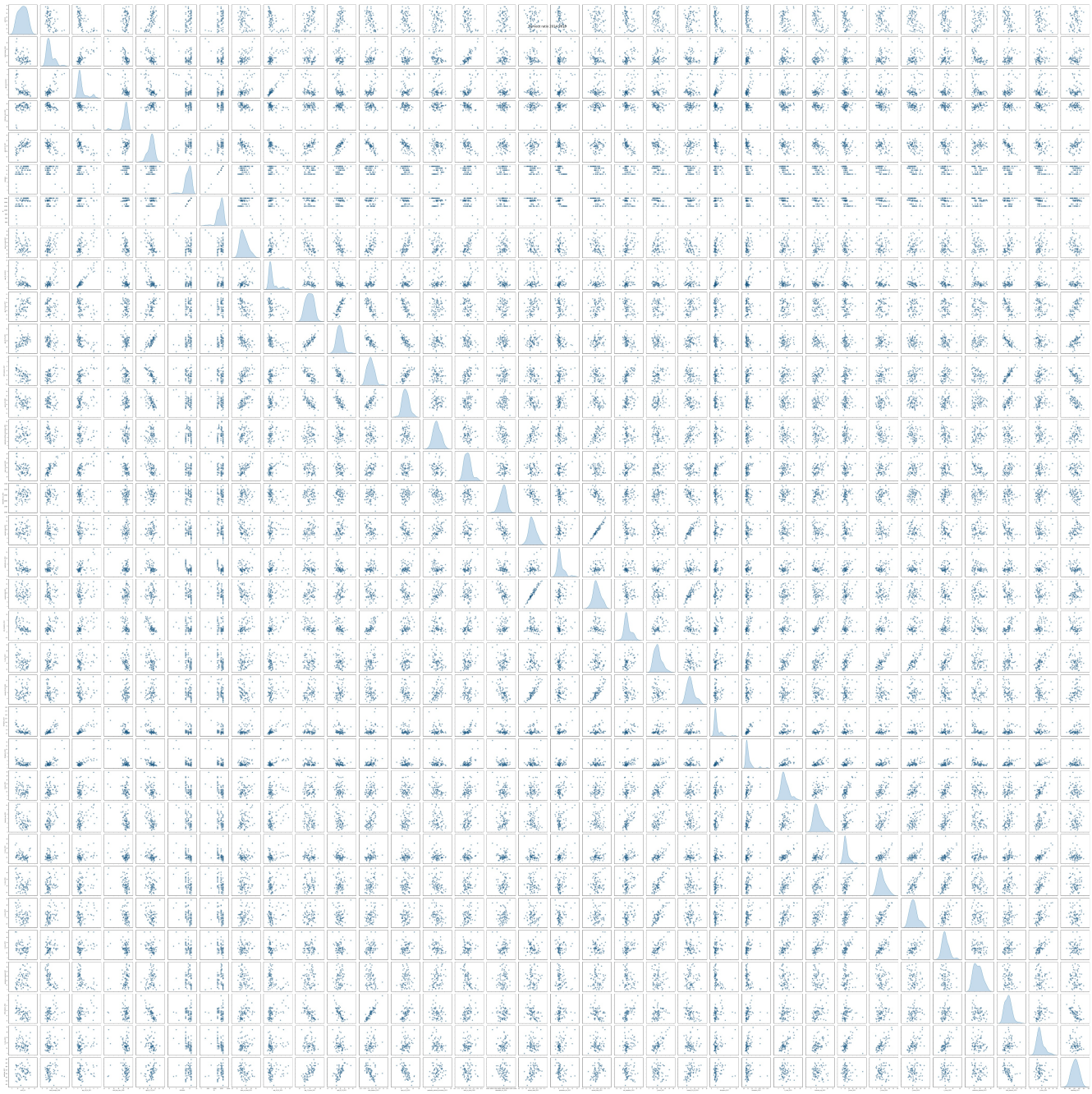
<seaborn.axisgrid.PairGrid at 0x298af872708>





In [21]:

```
# Plot colored by continent for years 2000-2007
sns.pairplot(df2[df2['vot19_14'] >= 7.9],  diag_kind = 'kde',
             plot_kws = {'alpha': 0.6, 's': 80, 'edgecolor': 'k'},
             size = 4);
# Title
plt.suptitle('Turnout ratio 2014-2019',
             size = 28);
```

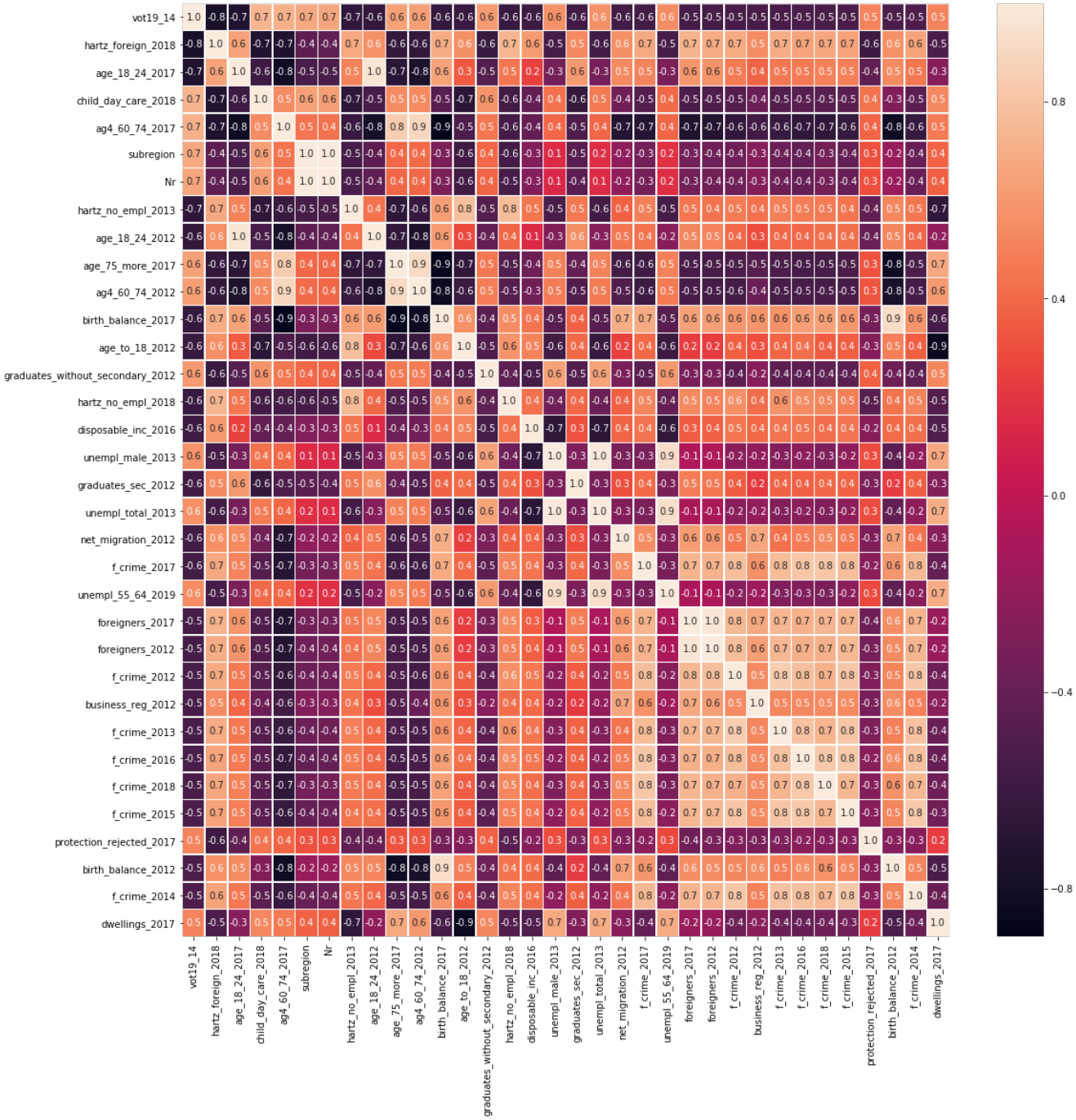


In [178]:

```
f,ax = plt.subplots(figsize=(18, 18))
sns.heatmap(df2.corr(), annot=True, linewidths=.5, fmt= '.1f',ax=ax)
```

Out[178]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1f5f86634a8>
```



- As the linear regression has the tendency to get overfitted with data having collinearity between independent varaibles, while fitting the data to a model we need to deal with multicolinearity between them .
- So in the forth coming feature selection and extraction steps we will deal with that using some feature selection techniques.
- From the above analysis with respect to target Region,state abbrev have been eliminated by the corr matrix as they dont really constitute to the target variable so be remove them from the data set before building our model

In [9]:

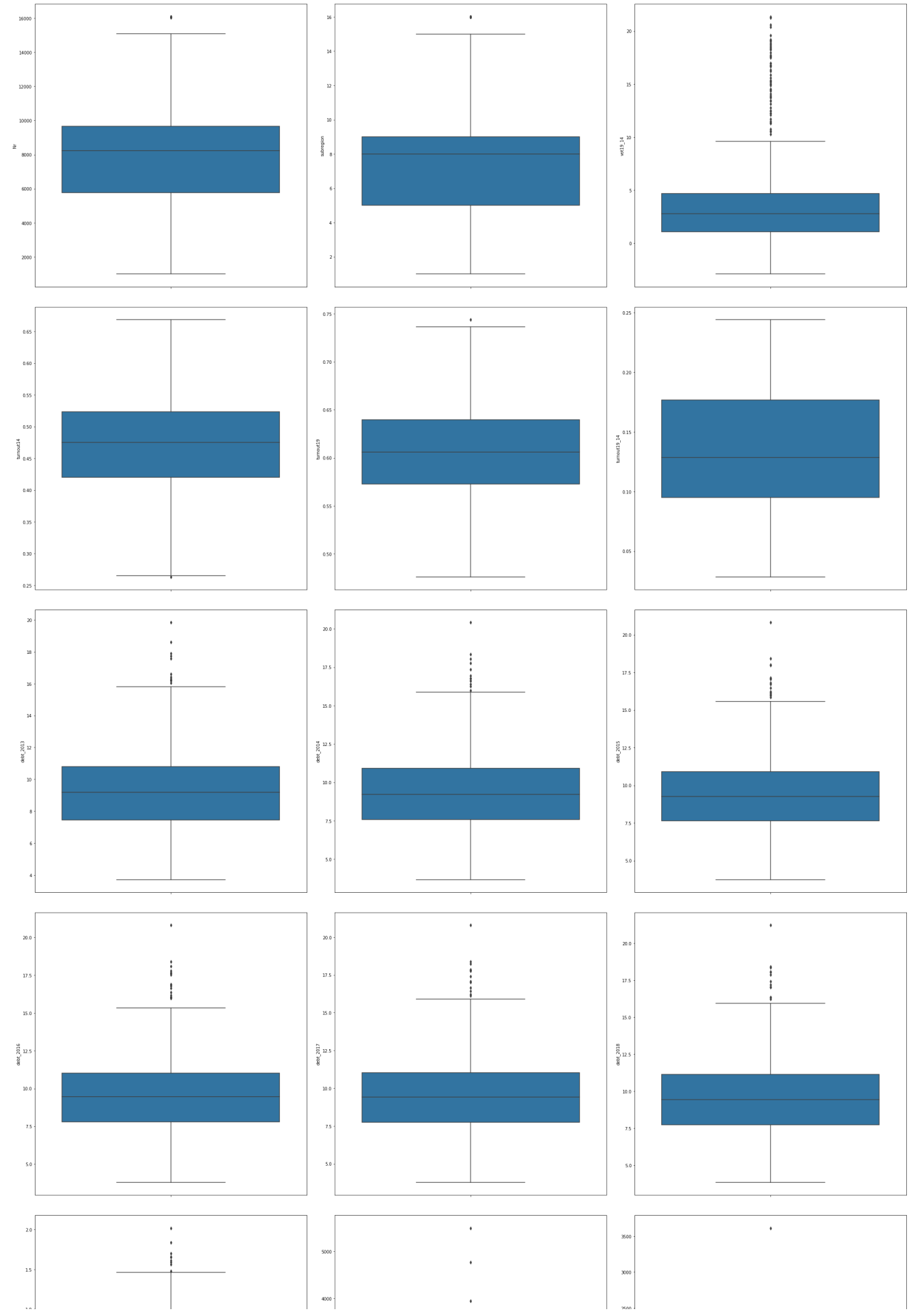
```
df1.drop(columns = {'region','state_abbrev','state'},axis = 1,inplace = True)
```

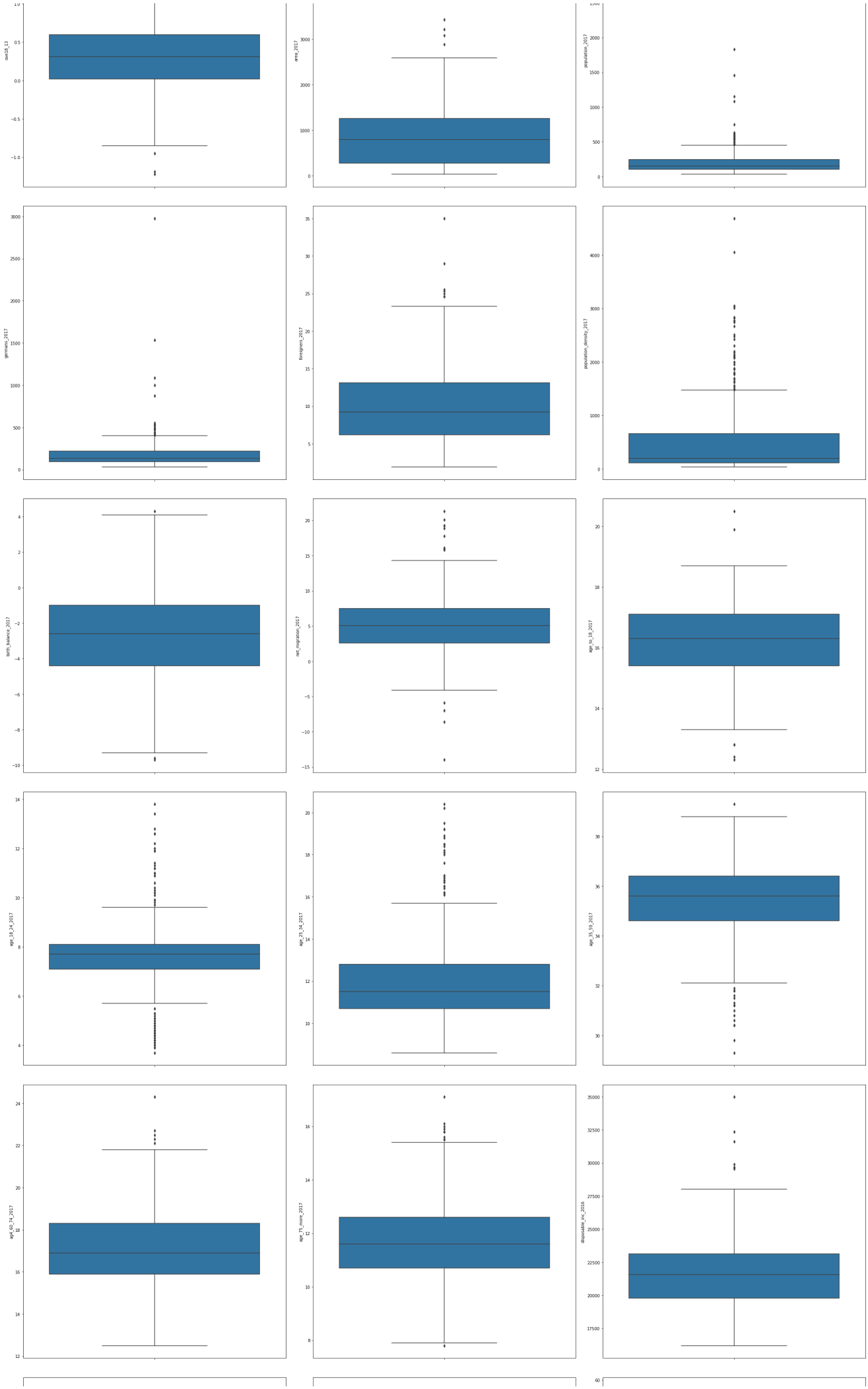
Lets visualize outliers using boxplots and find the no"s in the data

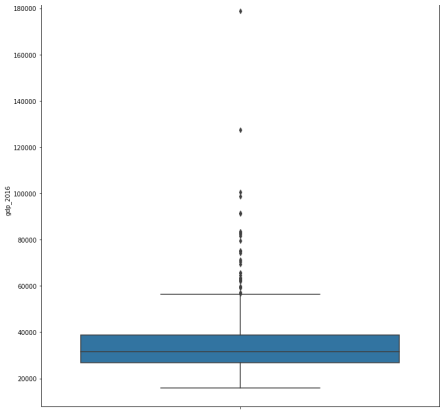
In [185]:

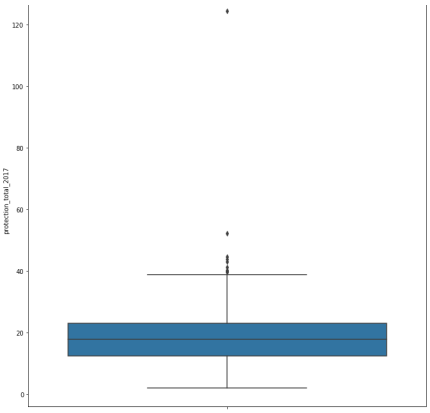
```
# from the above we find diff in min and max values so lets visualize the oultiers uisng Box plot
of  all the Attributes.
n_columns = 3
n_rows = 55
```

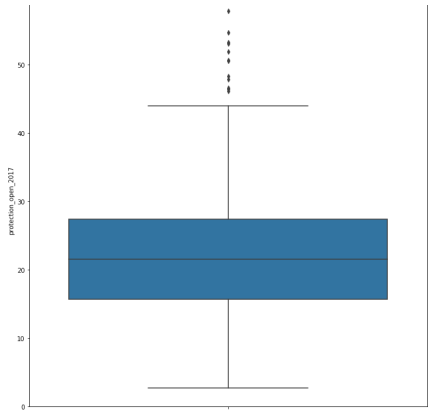
```
f, axes = plt.subplots(n_rows, n_columns, figsize=(10* n_columns, 10* n_rows))
for i, c in enumerate(df1.columns):
    sns.boxplot(y = c, data = df1, ax = axes[i // n_columns, i % n_columns])
plt.tight_layout()
plt.show()
```

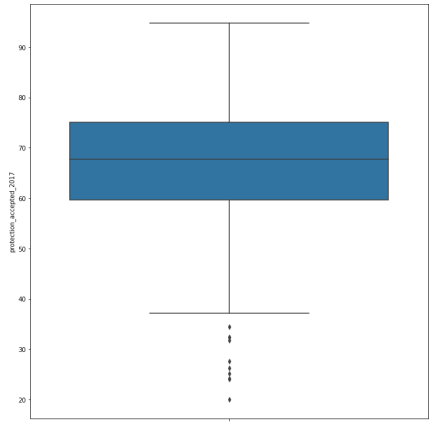


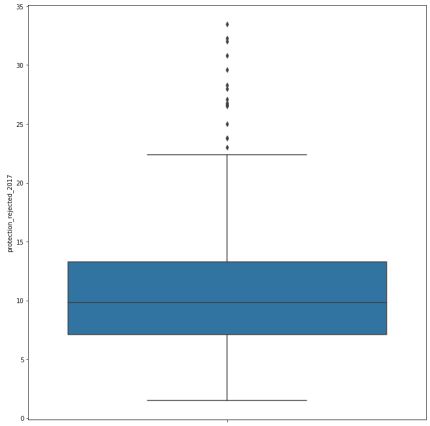

ovc18_13
area_2017
population_2017
germans_2017
foreigners_2017
population_density_2017
birth_balance_2017
net_migration_2017
age_to_18_2017
age_18_24_2017
age_25_34_2017
age_35_59_2017
age_60_74_2017
age_75_more_2017
disposable_inc_2016

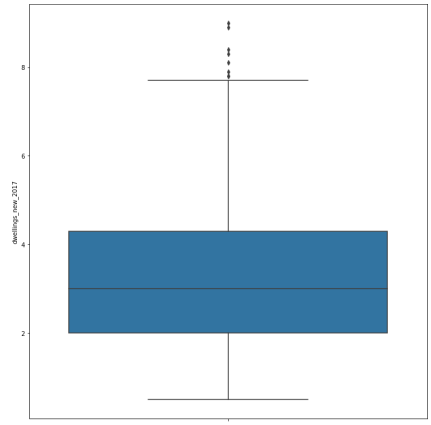

gdp_2016
180000
160000
140000
120000
100000
80000
60000
40000
20000

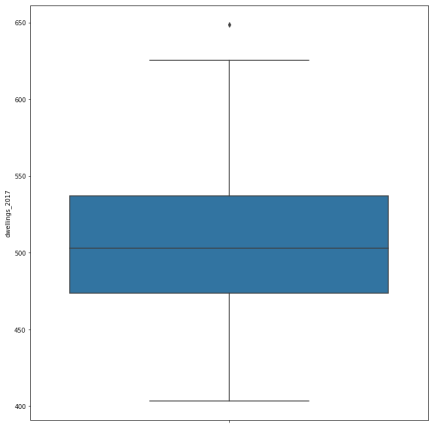

protection_total_2017
120
100
80
60
40
20
0

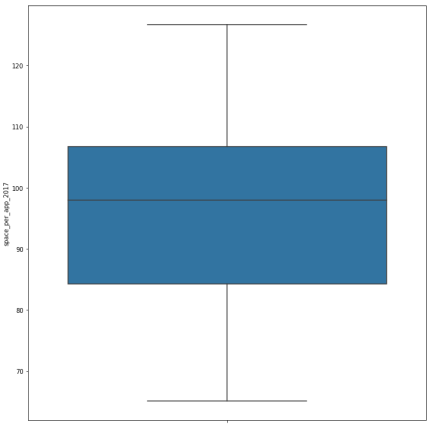

protection_open_2017
50
40
30
20
10
0

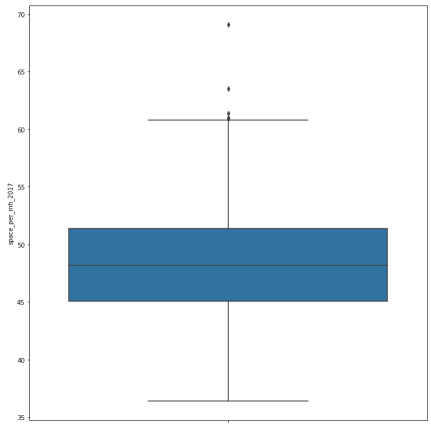

protection_accepted_2017
90
80
70
60
50
40
30
20

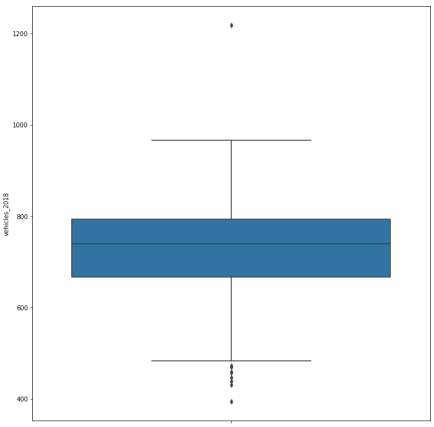

protection_rejected_2017
35
30
25
20
15
10
5
0

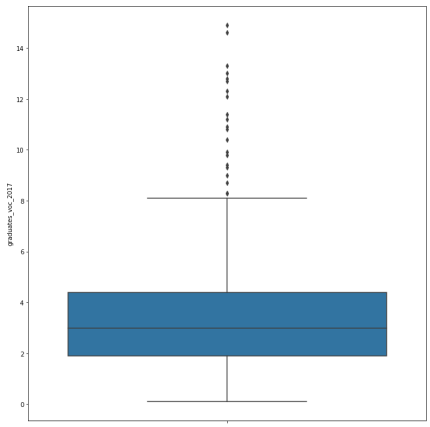

dwellings_new_2017
8
6
4
2

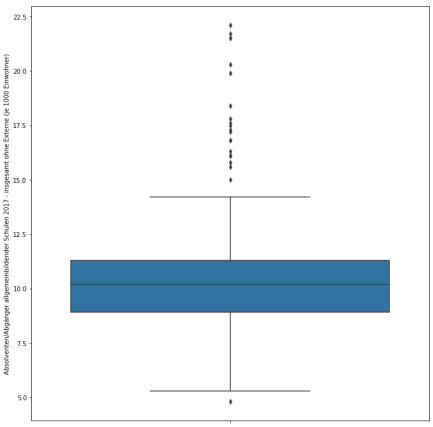

dwellings_2017
650
600
550
500
450
400

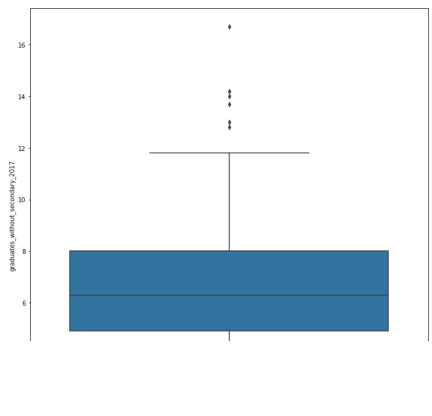

space_per_app_2017
120
110
100
90
80
70

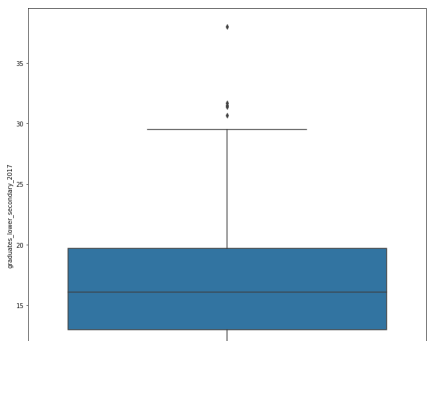

space_per_inh_2017
70
65
60
55
50
45
40
35

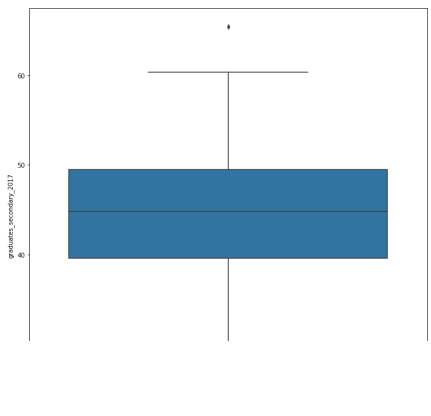

vehicles_2018
1200
1000
800
600
400

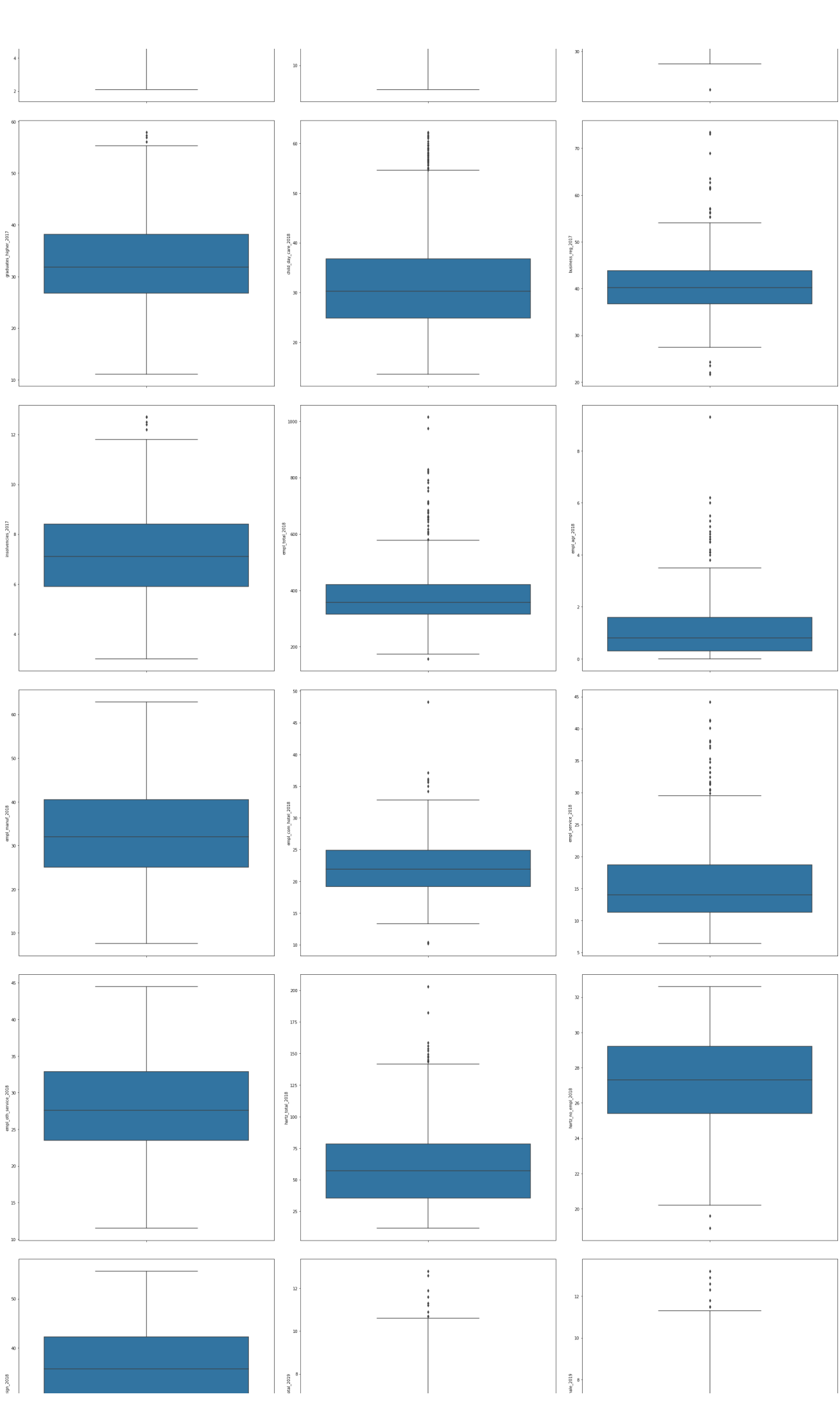

graduates_voc_2017
14
12
10
8
6
4
2
0

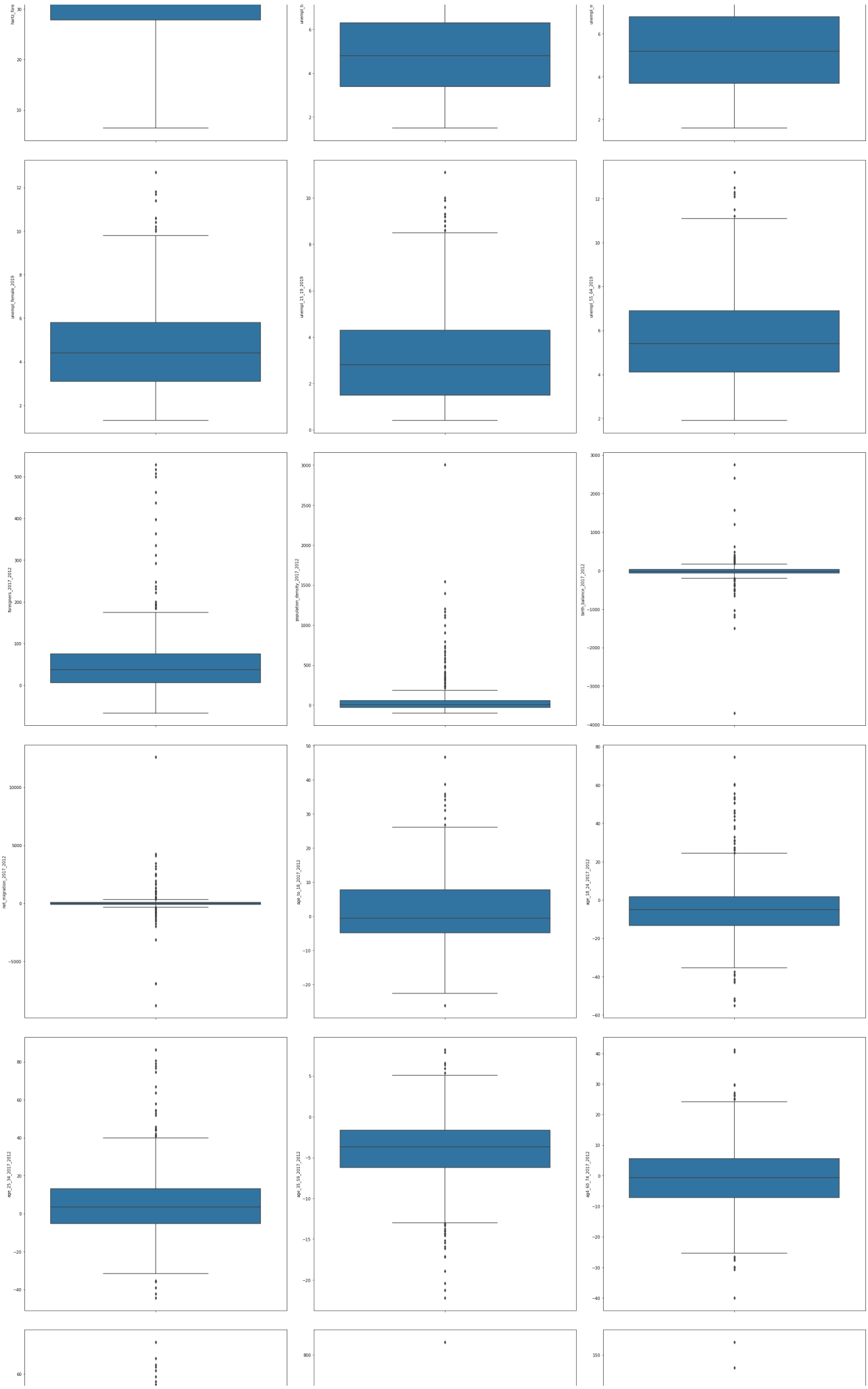

Absolventen/Abgänger allgemeinbildender Schulen 2017 - insgesamt ohne Externe (je 1000 Einwohner)
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0

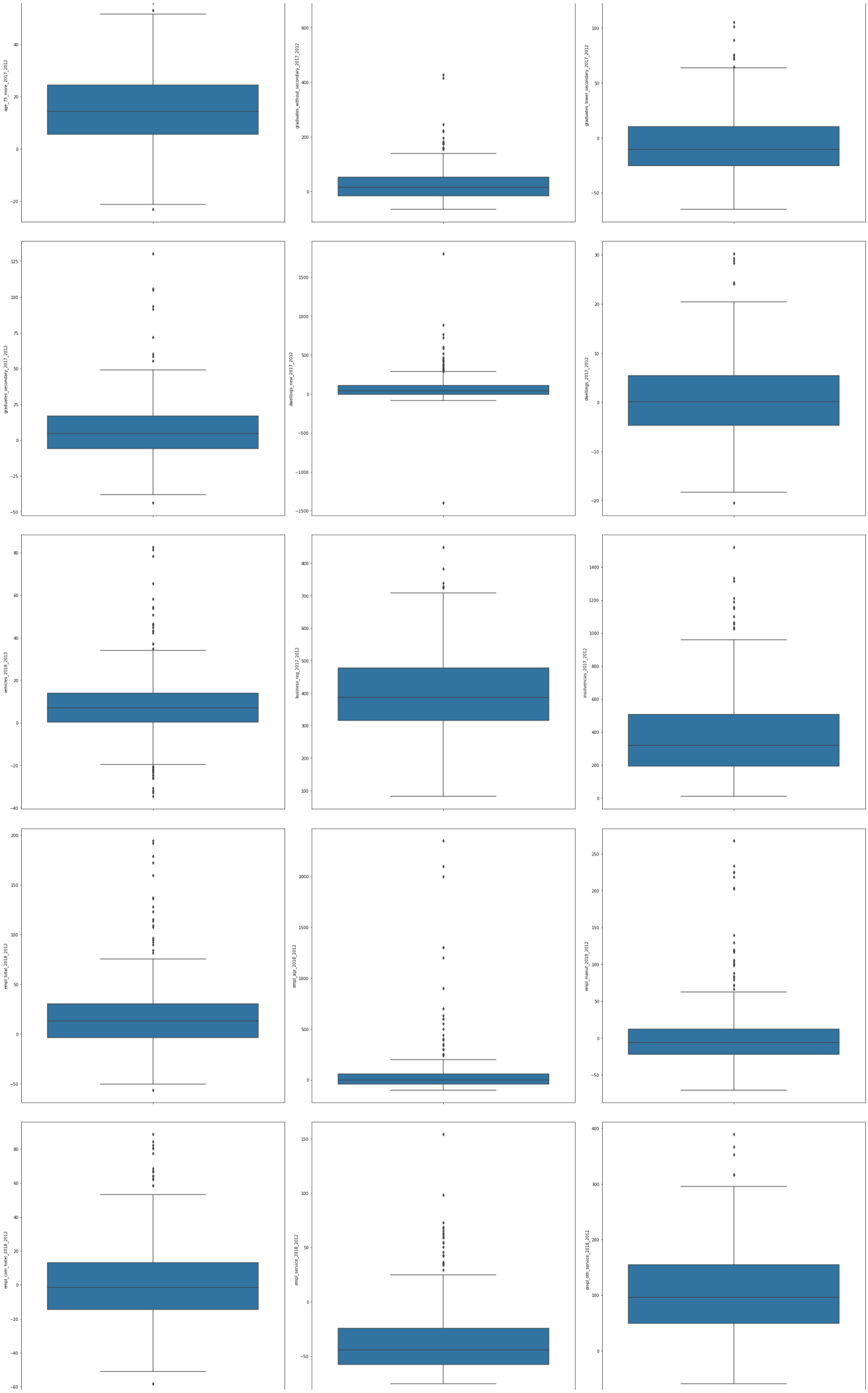

graduates_without_secondary_2017
16
14
12
10
8
6

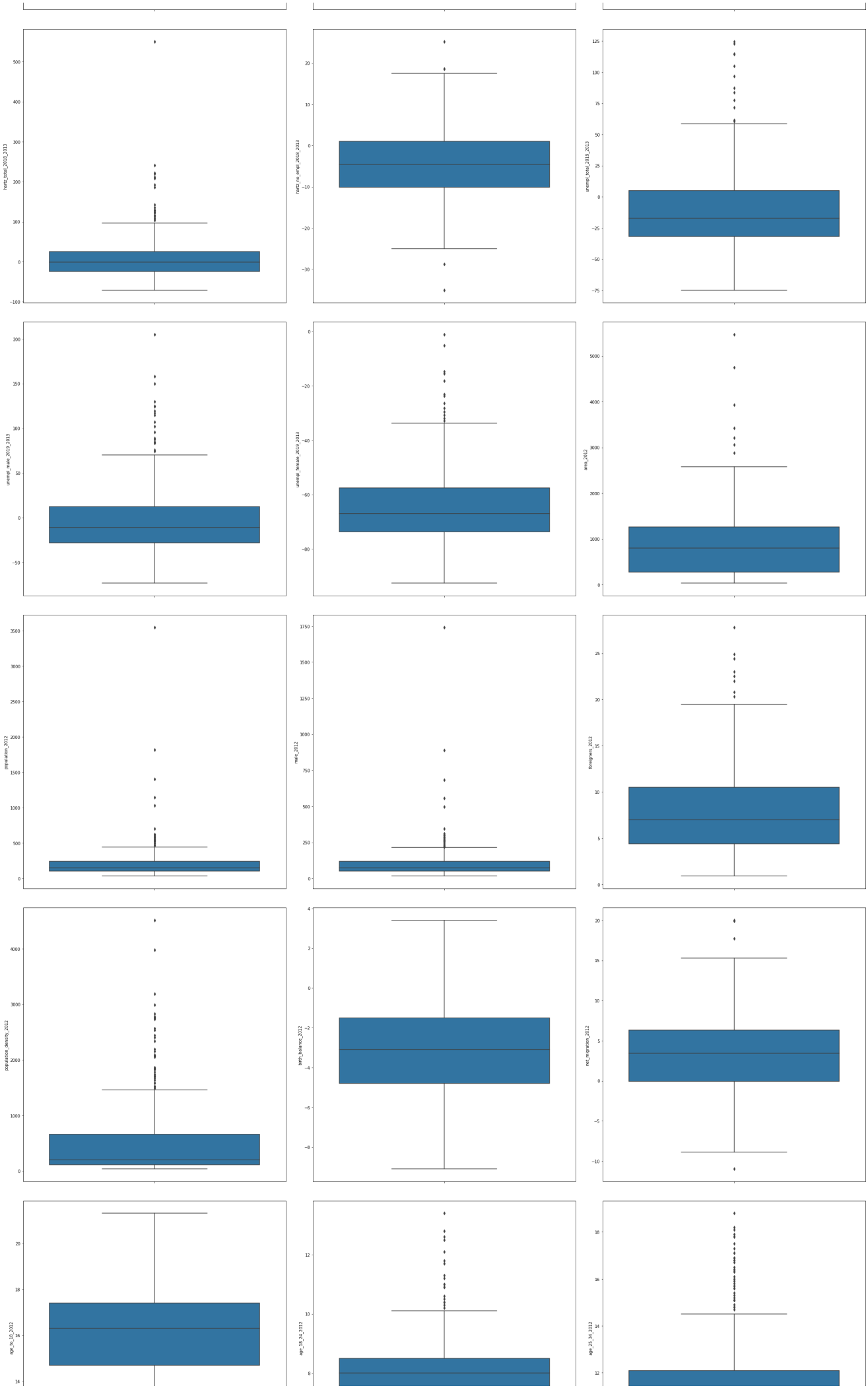

graduates_lower_secondary_2017
35
30
25
20
15

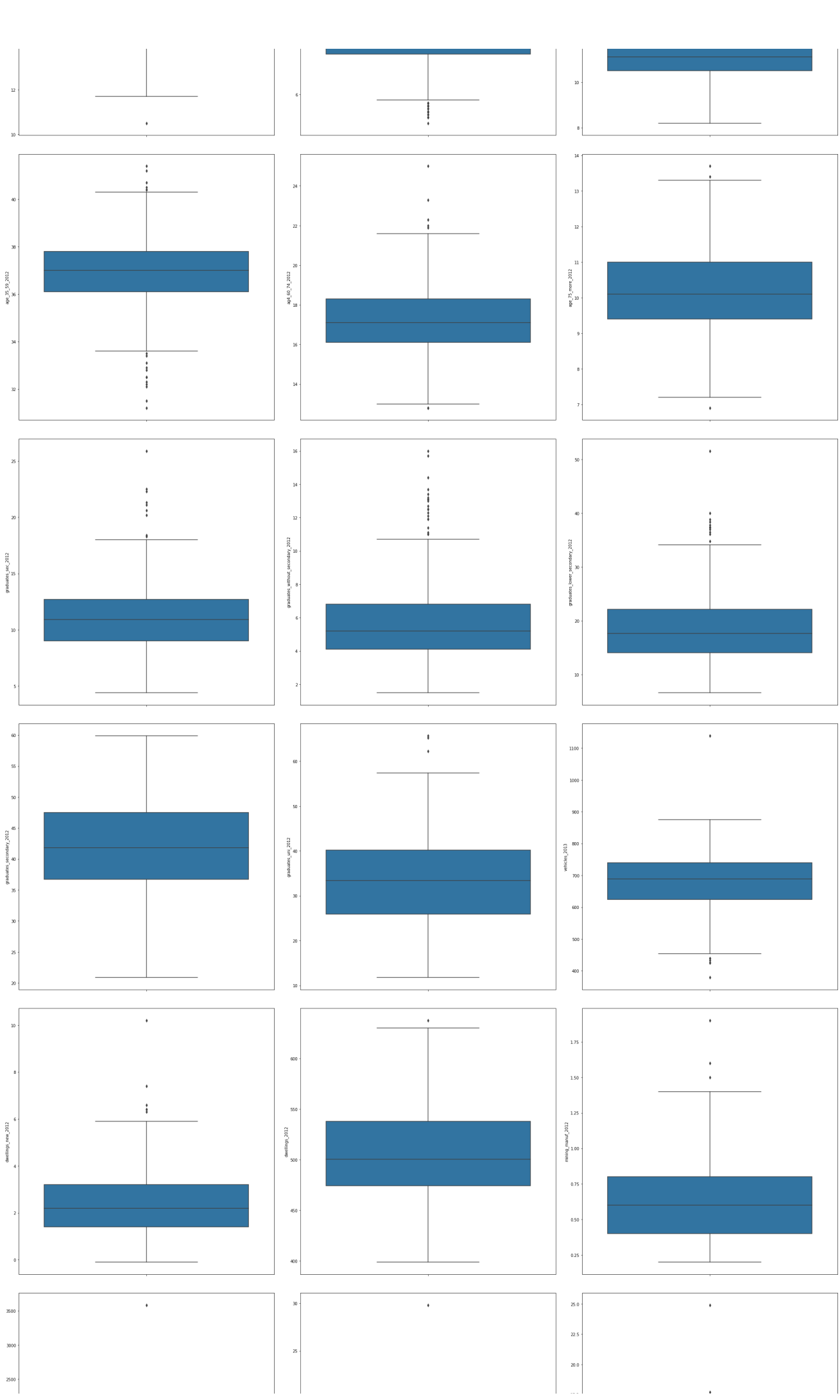

graduates_secondary_2017
60
50
40

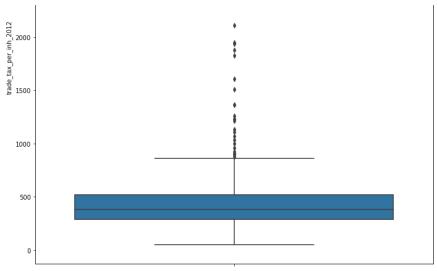

graduate_higher_2017
child_day_care_2018
business_mng_2017
insolvencies_2017
empl_total_2018
empl_agr_2018
empl_manuf_2018
empl_com_hotel_2018
empl_service_2018
empl_oth_service_2018
hunts_total_2018
hunts_no_empl_2018

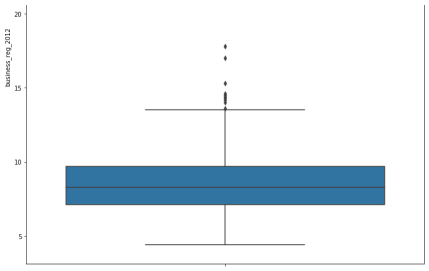

harts_fore
unempl_fe
unempl_fe
unempl_female_2019
unempl_15_19_2019
unempl_55_64_2019
foreigners_2017_2012
population_density_2017_2012
birth_balance_2017_2012
net_migration_2017_2012
age_0o_19_2017_2012
age_18_24_2017_2012
age_25_34_2017_2012
age_35_59_2017_2012
age_60_74_2017_2012

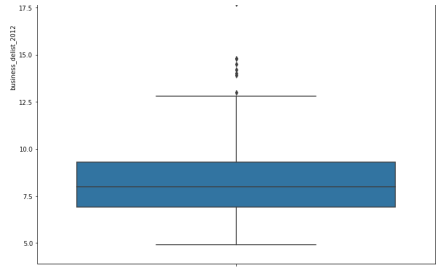

age_15_more_2017_2012
graduates_without_secondary_2017_2012
graduates_lower_secondary_2017_2012
graduates_secondary_2017_2012
dwellings_new_2017_2012
dwellings_2017_2012
vehicles_2018_2013
business_reg_2017_2012
insolvencies_2017_2012
empl_total_2018_2012
empl_agr_2018_2012
empl_manuf_2018_2012
empl_com_hotel_2018_2012
empl_service_2018_2012
empl_oth_service_2018_2012

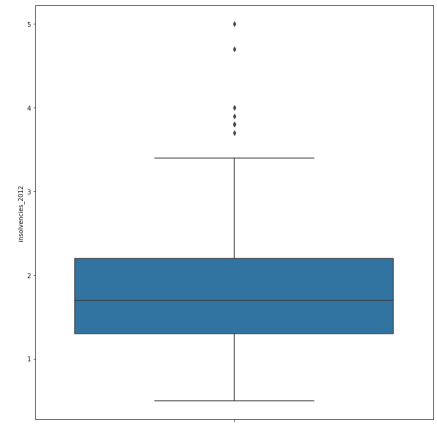

hartz_total_2018_2013
hartz_no_empl_2018_2013
unempl_total_2019_2013
unempl_male_2019_2013
unempl_female_2019_2013
area_2012
population_2012
male_2012
foreigners_2012
population_density_2012
birth_balance_2012
net_migration_2012
age_to_18_2012
age_18_24_2012
age_25_34_2012

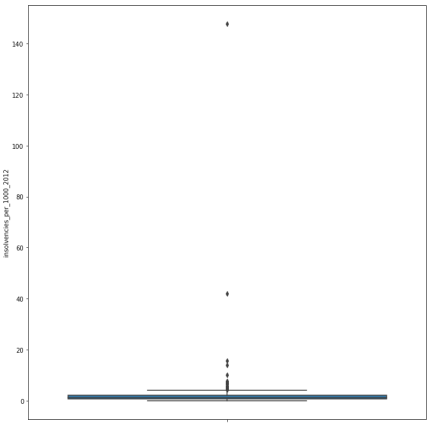

age_35_59_2012
age_60_74_2012
age_75_more_2012
graduates_sec_2012
graduates_without_secondary_2012
graduates_lower_secondary_2012
graduates_secondary_2012
graduates_uni_2012
vehicles_2013
dwellings_new_2012
dwellings_2012
mining_manuf_2012

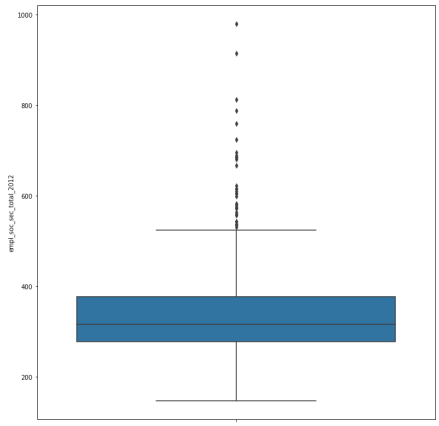

trade_tax_per_oth_2012

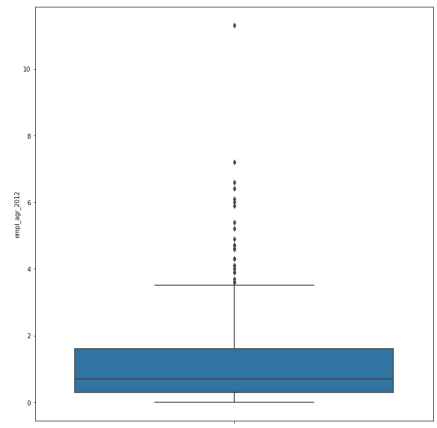

business_reg_2012

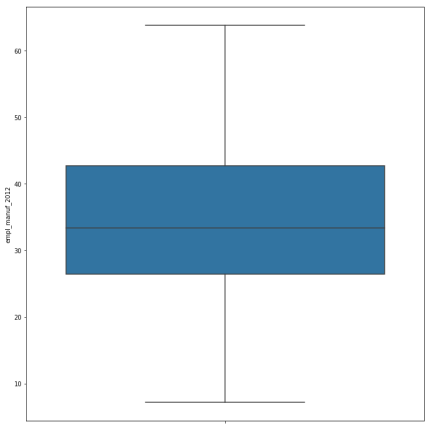

business_dereg_2012

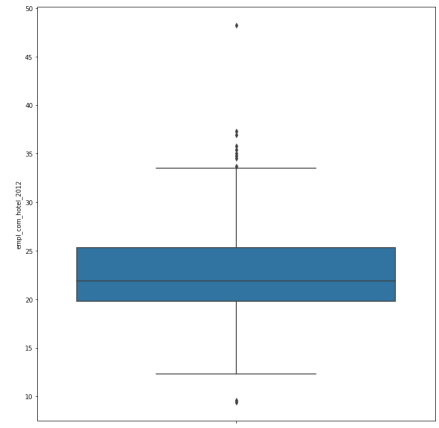

insolvencies_2012

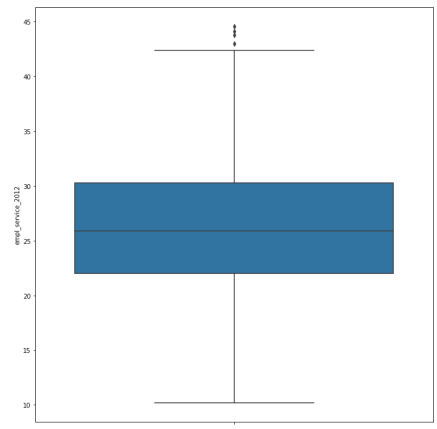

insolvencies_per_1000_2012

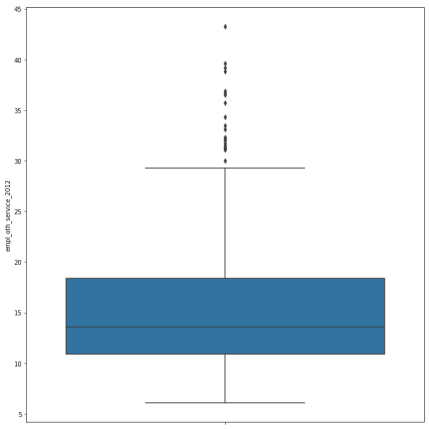

empl_soc_sec_total_2012

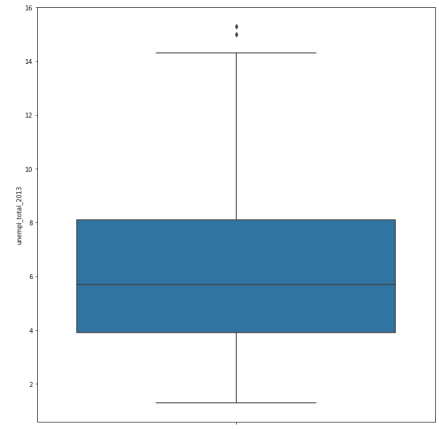

empl_agr_2012

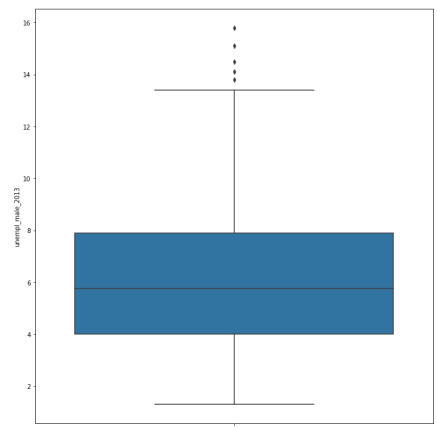

empl_manuf_2012

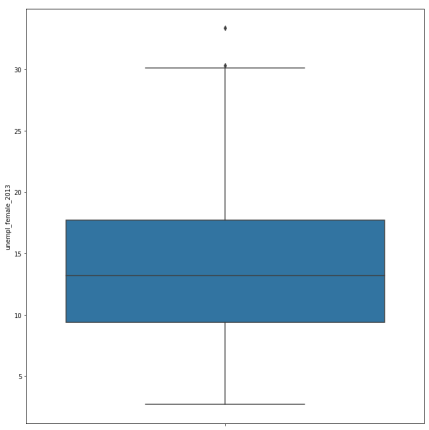

empl_com_hotel_2012

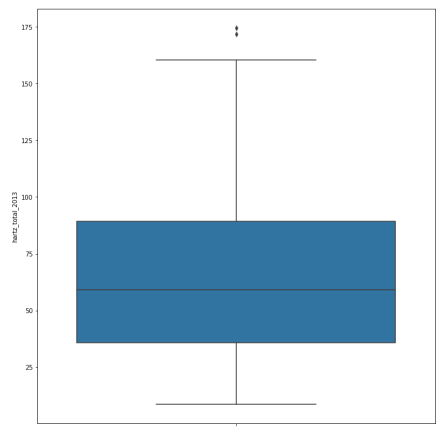

empl_service_2012

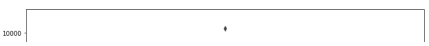

empl_oth_service_2012

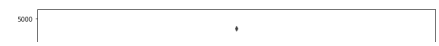

unempl_total_2013

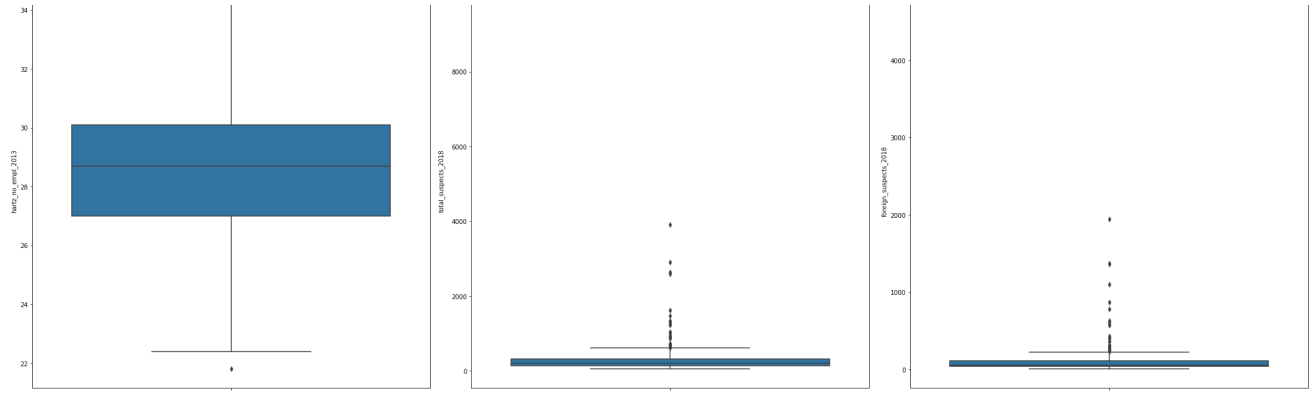

unempl_male_2013


unempl_female_2013


hartz_total_2013

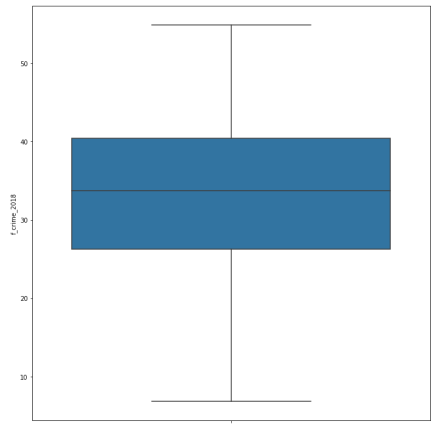


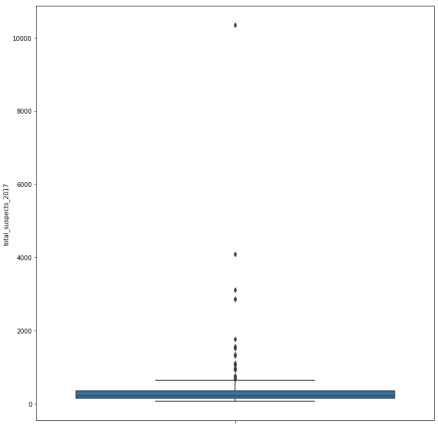


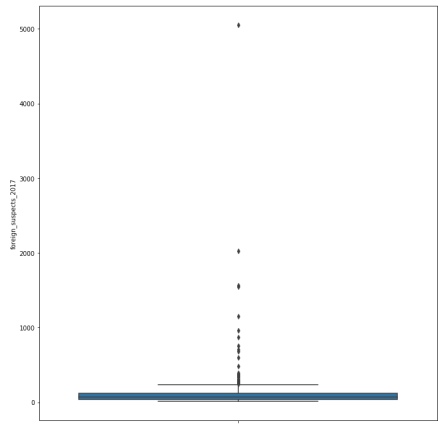

hwtz_no_empl_2013

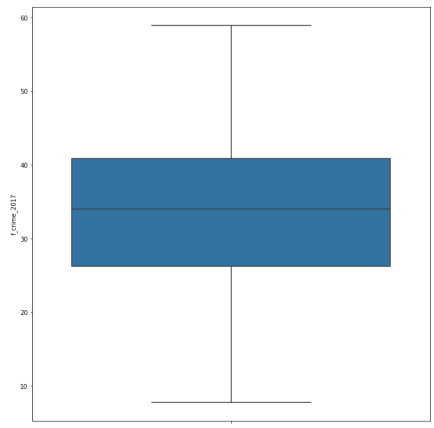
total_suspects_2018

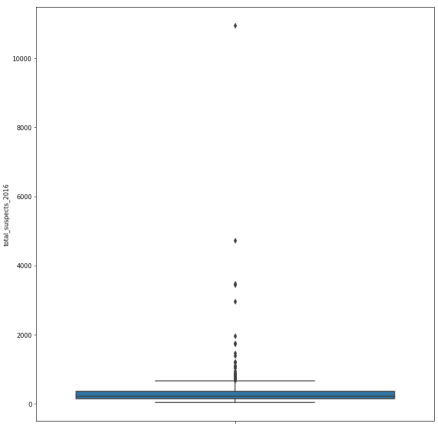
foreign_suspects_2018

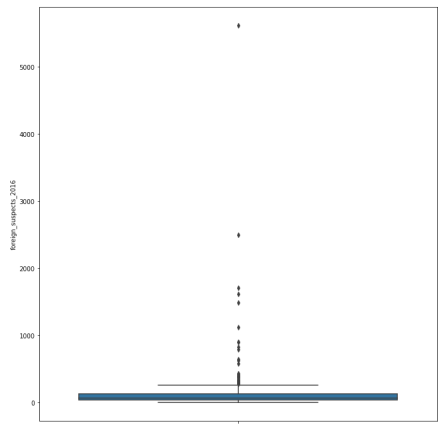

f_crime_2018

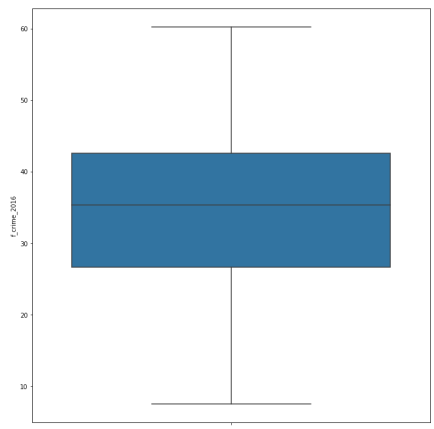

total_suspects_2017

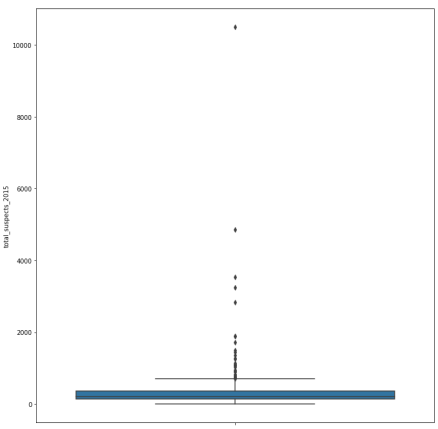

foreign_suspects_2017

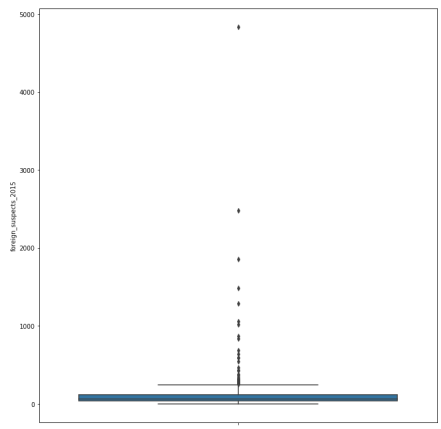

f_crime_2017

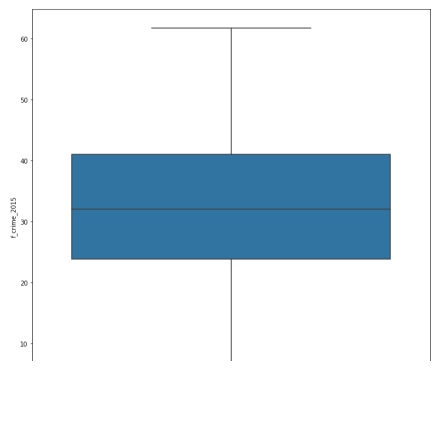

total_suspects_2016

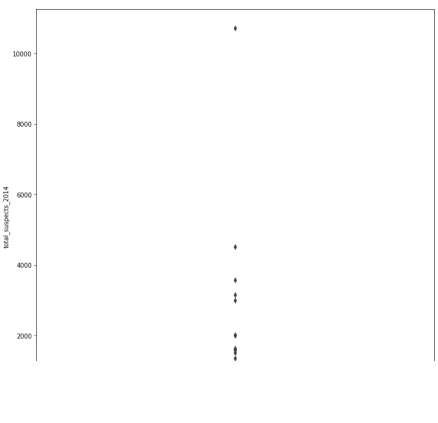

foreign_suspects_2016

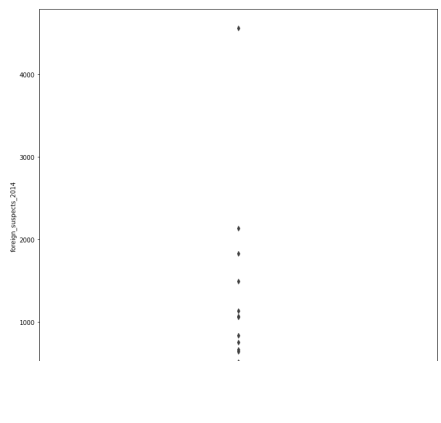

f_crime_2016

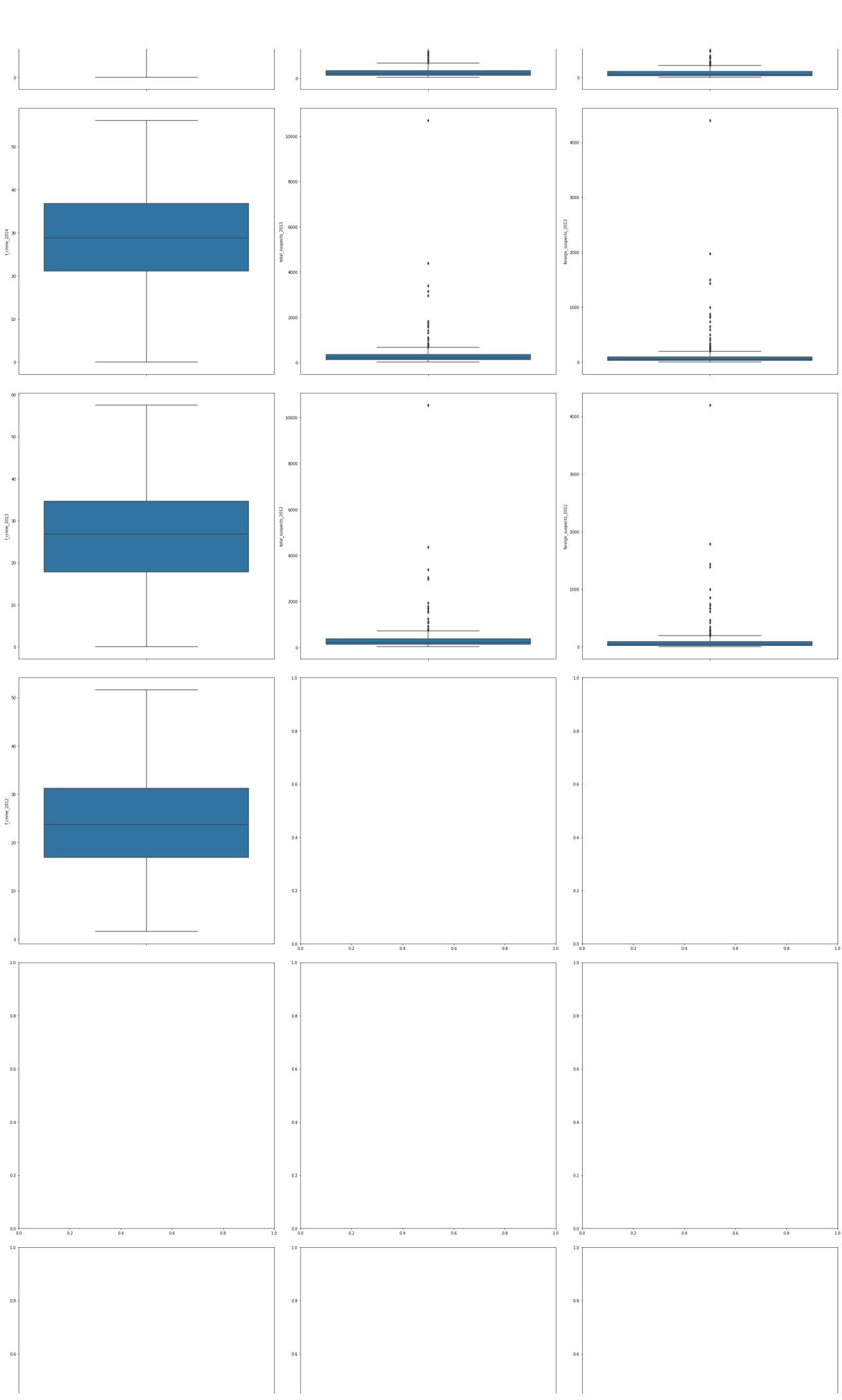

total_suspects_2015


foreign_suspects_2015


f_crime_2015


total_suspects_2014


foreign_suspects_2014


f_crime_2014
total_suspects_2013
foreign_suspects_2013
f_crime_2013
total_suspects_2012
foreign_suspects_2012
f_crime_2012

- As already seen max no of attributes have outliers in the data so while fitting the data these parameters shouold be taken care of as the LG overcompensates for outliers.

In [186]:

```
# we can even find them using the following code
def remove_outlier(df1):
    for i in df1.columns:
        q1 = df1[i].quantile(0.25)
        q3 = df1[i].quantile(0.75)
        iqr = q3-q1 #Interquartile range
        item1_low  = q1-1.5*iqr
        item1_high = q3+1.5*iqr
        df_out = df1.loc[(df1[i] < item1_low) | (df1[i] > item1_high)]
        return df_out
```

In [187]:

```
remove_outlier(df1)["vot19_14"] # the below are the ouliers in the data
```

Out[187]:

```
378     9.566501
379    18.453636
380     5.929222
381    10.577550
382     8.346655
383    14.983195
384    12.764963
385    16.782858
386    16.664091
387    16.973704
388    18.334998
389    15.582975
390    15.260989
391    17.965880
392    17.625163
393    15.364231
394    16.939578
395    19.574445
396    18.712021
397    13.389590
398    18.719615
399    15.881514
400    19.110053
Name: vot19_14, dtype: float64
```

feature selection and model building

model1 : feature selection using only features that are not highly colinear to each other

lets just calculate the corr of the fatures having r > 0.8 , Statistically when r>0.8 then the two varaibles or various variable are said to have multi collinearity between each other as they tend to expalain each other very well i.e linealy changing.

In [75]:

```
df3 = df1.copy()
df3.head(2)
```

Out[75]:

	Nr	subregion	vot19_14	turnout14	turnout19	turnout19_14	debt_2013	debt_2014	debt_2015	debt_2016	debt_2017	debt_2018
0	1001	1.0	-0.003027	0.357400	0.562920	0.205520	16.41	16.40	16.21	16.17	16.21	16.24
1	1002	1.0	0.060316	0.402589	0.588603	0.186015	12.04	12.03	12.17	12.23	12.16	11.96

In [76]:

```
from sklearn.model_selection import train_test_split
X= df3.drop("vot19_14", axis = 1)
Y = df3["vot19_14"]
```

In [77]:

```
corr = X.corr()
columns = np.full((corr.shape[0],), True, dtype=bool)
for i in range(corr.shape[0]):
    for j in range(i+1, corr.shape[0]):
        if corr.iloc[i,j] >= 0.8:
```

```
            if columns[j]:
                columns[j] = False
```

In [78]:

```
corelated_columns = X.columns[columns]
corelated_columns.shape    # these are highly correlated features which have to be removed
```

Out[78]:

```
(78,)
```

In [79]:

```
X.drop(X.columns[columns], axis = 1, inplace = True)
```

In [80]:

```
### moded1 : using stats model why eliminating the correlated features from the data frame


x1 = sm.add_constant(X) #Adding the constant
model = sm.OLS(Y,x1).fit() # fitting the model
print(model.summary()) # model summary


def checkVIF(x):
    vif = pd.DataFrame()
    vif['Features'] = x.columns
    vif['VIF'] = [variance_inflation_factor(x.values, i) for i in range(x.shape[1])]
    vif['VIF'] = round(vif['VIF'], 2)
    vif = vif.sort_values(by = "VIF", ascending = False)
    return(vif)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                vot19_14   R-squared:                       0.902
Model:                            OLS   Adj. R-squared:                  0.882
Method:                 Least Squares   F-statistic:                     44.32
Date:                Wed, 29 Jan 2020   Prob (F-statistic):          2.30e-131
Time:                        22:13:26   Log-Likelihood:                -771.13
No. Observations:                 401   AIC:                             1682.
Df Residuals:                     331   BIC:                             1962.
Df Model:                          69
Covariance Type:            nonrobust
================================================================================================
                                   coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------
const                          117.5092    209.038      0.562      0.574    -293.701      528.719
subregion                        0.4018      0.060      6.664      0.000       0.283        0.520
debt_2014                       -0.9748      0.654     -1.491      0.137      -2.261        0.311
debt_2015                       -0.0712      0.817     -0.087      0.931      -1.679        1.536
debt_2016                        0.6394      1.171      0.546      0.585      -1.664        2.943
debt_2017                       -0.9414      1.373     -0.686      0.493      -3.643        1.760
debt_2018                        1.4874      0.881      1.688      0.092      -0.246        3.221
germans_2017                    -0.0123      0.020     -0.627      0.531      -0.051        0.026
age_75_more_2017                 0.1428      0.355      0.402      0.688      -0.556        0.842
space_per_inh_2017              -0.0653      0.050     -1.318      0.188      -0.163        0.032
empl_total_2018                 -0.0025      0.008     -0.332      0.740      -0.017        0.012
hartz_total_2018                 0.0255      0.019      1.319      0.188      -0.013        0.063
unempl_total_2019                2.3458      2.868      0.818      0.414      -3.295        7.987
unempl_male_2019                -2.0287      1.590     -1.276      0.203      -5.156        1.099
unempl_female_2019              -0.5468      1.343     -0.407      0.684      -3.188        2.094
unempl_15_19_2019               -0.1701      0.125     -1.366      0.173      -0.415        0.075
unempl_55_64_2019                0.7541      0.213      3.543      0.000       0.335        1.173
age_25_34_2017_2012             -0.0051      0.008     -0.643      0.520      -0.021        0.011
empl_service_2018_2012           0.0077      0.006      1.225      0.222      -0.005        0.020
unempl_total_2019_2013          -0.0286      0.019     -1.474      0.141      -0.067        0.010
unempl_male_2019_2013            0.0242      0.017      1.423      0.156      -0.009        0.058
area_2012                       -0.0005      0.000     -1.821      0.069      -0.001     3.79e-05
population_2012                 -0.0497      0.049     -1.006      0.315      -0.147        0.048
male_2012                        0.1316      0.091      1.440      0.151      -0.048        0.311
foreigners_2012                 -0.2712      0.072     -3.758      0.000      -0.413       -0.129
population_density_2012         -0.0011      0.000     -2.626      0.009      -0.002       -0.000
```

population_density_2012	-0.0011	0.000	-2.626	0.009	-0.002	-0.000
birth_balance_2012	-0.1974	0.164	-1.206	0.229	-0.519	0.125
age_to_18_2012	-0.5630	1.423	-0.396	0.693	-3.363	2.237
age_18_24_2012	-1.5267	1.414	-1.080	0.281	-4.307	1.254
age_25_34_2012	0.4041	1.436	0.281	0.779	-2.421	3.229
age_35_59_2012	-0.2232	1.410	-0.158	0.874	-2.997	2.550
ag4_60_74_2012	-0.7916	1.410	-0.561	0.575	-3.566	1.982
age_75_more_2012	-0.4221	1.429	-0.295	0.768	-3.233	2.389
graduates_sec_2012	-0.1359	0.067	-2.029	0.043	-0.268	-0.004
graduates_lower_secondary_2012	-0.1273	0.026	-4.942	0.000	-0.178	-0.077
graduates_uni_2012	-0.0179	0.019	-0.942	0.347	-0.055	0.019
vehicles_2013	0.0010	0.003	0.360	0.719	-0.004	0.006
dwellings_2012	-0.0024	0.006	-0.379	0.705	-0.015	0.010
trade_tax_per_inh_2012	0.0002	0.001	0.286	0.775	-0.001	0.001
business_delist_2012	-0.1641	0.076	-2.152	0.032	-0.314	-0.014
empl_soc_sec_total_2012	-0.0006	0.008	-0.069	0.945	-0.017	0.015
empl_agr_2012	-0.5554	1.544	-0.360	0.719	-3.593	2.483
empl_manuf_2012	-0.6388	1.539	-0.415	0.678	-3.667	2.389
empl_com_hotel_2012	-0.7250	1.538	-0.471	0.638	-3.751	2.301
empl_service_2012	-0.6750	1.539	-0.438	0.661	-3.703	2.353
empl_oth_service_2012	-0.7842	1.543	-0.508	0.612	-3.819	2.250
unempl_total_2013	-1.1879	0.500	-2.374	0.018	-2.172	-0.204
unempl_male_2013	1.1205	0.468	2.393	0.017	0.200	2.041
unempl_female_2013	-0.0073	0.046	-0.158	0.874	-0.098	0.084
hartz_total_2013	0.0259	0.022	1.198	0.232	-0.017	0.069
total_suspects_2018	-0.0028	0.004	-0.682	0.495	-0.011	0.005
foreign_suspects_2018	0.0126	0.009	1.457	0.146	-0.004	0.030
total_suspects_2017	0.0010	0.005	0.187	0.852	-0.009	0.011
foreign_suspects_2017	-0.0082	0.011	-0.750	0.454	-0.030	0.013
f_crime_2017	0.0123	0.028	0.440	0.660	-0.043	0.067
total_suspects_2016	-0.0038	0.005	-0.804	0.422	-0.013	0.006
foreign_suspects_2016	-0.0007	0.009	-0.075	0.940	-0.019	0.017
f_crime_2016	-0.0168	0.028	-0.596	0.551	-0.072	0.039
total_suspects_2015	-0.0008	0.005	-0.150	0.881	-0.011	0.009
foreign_suspects_2015	0.0092	0.010	0.913	0.362	-0.011	0.029
f_crime_2015	0.0025	0.026	0.096	0.924	-0.049	0.054
total_suspects_2014	-0.0015	0.003	-0.560	0.576	-0.007	0.004
foreign_suspects_2014	0.0012	0.008	0.149	0.882	-0.015	0.017
f_crime_2014	0.0101	0.025	0.407	0.685	-0.039	0.059
total_suspects_2013	0.0033	0.004	0.806	0.421	-0.005	0.011
foreign_suspects_2013	-0.0111	0.010	-1.145	0.253	-0.030	0.008
f_crime_2013	0.0071	0.026	0.275	0.784	-0.044	0.058
total_suspects_2012	0.0020	0.004	0.476	0.635	-0.006	0.010
foreign_suspects_2012	-0.0023	0.011	-0.216	0.829	-0.023	0.019
f_crime_2012	-0.0159	0.030	-0.529	0.597	-0.075	0.043

Omnibus:	11.959	Durbin-Watson:	1.721
Prob(Omnibus):	0.003	Jarque-Bera (JB):	15.978
Skew:	0.266	Prob(JB):	0.000339
Kurtosis:	3.821	Cond. No.	5.49e+06

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 5.49e+06. This might indicate that there are
strong multicollinearity or other numerical problems.

```
C:\Users\SOURAV CH\Anaconda3\lib\site-packages\numpy\core\fromnumeric.py:2389: FutureWarning: Meth
od .ptp is deprecated and will be removed in a future version. Use numpy.ptp instead.
  return ptp(axis=axis, out=out, **kwargs)
```

Durbin Watson Statistic (DW) - This test is used to check autocorrelation. Its value lies between 0 and 4. A DW=2 value shows no autocorrelation. However, a value between 0 > DW >=1 implies positive autocorrelation, while 2 < DW < 4 implies negative autocorrelation.. Variance Inflation Factor (VIF) - This metric is used to check multicollinearity. VIF <=4 implies no multicollinearity but VIF >=10 suggests high multicollinearity. Alternatively, you can also look at the tolerance (1/VIF) value to determine correlation in IVs. In addition, you can also create a correlation matrix to determine collinear variables.. Breusch-Pagan / Cook Weisberg Test - This test is used to determine presence of heteroskedasticity. If you find p < 0.05, you reject the null hypothesis and infer that heteroskedasticity is not present.

In [81]:

```
from scipy.stats import shapiro
shapiro(model.resid)
```

Out[81]:

```
(0.9914925694465637, 0.02114972658455372)
```

From the Shapiro results, since pvalue is less than 0.05 we can conclude that residuals do follow normal distribution

cooks distance

In [54]:

```
influence = model.get_influence()
cook = influence.summary_frame()
cook
```

Out[54]:

	dfb_const	dfb_subregion	dfb_debt_2014	dfb_debt_2015	dfb_debt_2016	dfb_debt_2017	dfb_debt_2018	dfb_germans_2017	dfb_a
0	0.042052	-0.038832	0.020715	0.012420	-0.051741	0.038788	-0.011636	-0.013273	
1	0.006215	0.026378	0.016933	-0.006022	-0.010303	0.001892	0.006012	-0.003712	
2	-0.032482	-0.067706	0.027350	-0.032855	0.010411	0.030986	-0.051268	-0.016464	
3	0.005221	-0.006651	-0.002768	0.002048	-0.009344	-0.001411	0.016436	-0.008095	
4	-0.057230	-0.064530	0.082622	-0.110199	-0.007166	0.023843	0.013816	-0.028787	
...	...	...	...	...	...	...	...	...	...
396	0.002100	0.042552	-0.051500	0.015809	0.027206	-0.029299	0.016533	0.018497	
397	0.000327	0.000063	-0.000673	0.000623	-0.000293	0.000361	-0.000194	0.000402	
398	0.095855	0.119944	0.163999	-0.088456	-0.037784	0.015610	-0.018229	-0.100427	
399	0.011189	-0.006101	0.000047	-0.000768	0.013067	-0.010805	-0.002243	0.007272	
400	-0.118078	0.049424	0.040498	-0.027325	0.013209	-0.045417	0.024348	-0.068116	

401 rows × 76 columns

In [55]:

```
cook[cook.cooks_d > 0.1]
```

Out[55]:

	dfb_const	dfb_subregion	dfb_debt_2014	dfb_debt_2015	dfb_debt_2016	dfb_debt_2017	dfb_debt_2018	dfb_germans_2017	dfb_a
66	0.016153	-0.146661	-0.085222	0.016375	0.144280	-0.132451	0.036901	0.084133	
324	-0.033384	0.130111	0.001108	0.010598	-0.025763	-0.007164	0.023979	0.409991	
395	-0.051812	-0.144855	0.521765	-0.221503	-0.303196	0.063467	0.184109	-0.017223	

For our data a feature with cooks distance d >0.1 is a outlier which is given above .

In [58]:

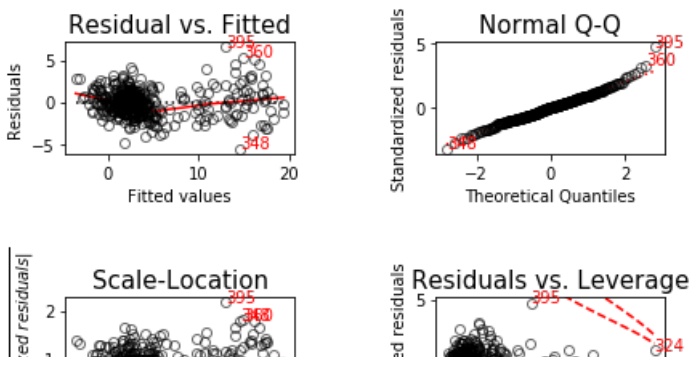
```
!pip install lmdiag
```

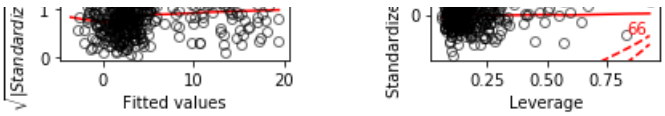
```
!pip install lmdiag
```

```
Collecting lmdiag
  Downloading
https://files.pythonhosted.org/packages/db/76/8b15c0e5065156fa776fe117877a18ff91f2218fb35d79d70b9b5
a59/lmdiag-0.3.7-py3-none-any.whl
Collecting linearmodels (from lmdiag)
  Downloading
https://files.pythonhosted.org/packages/dd/3f/ab18e5b2dc88be8e78974c5518813b0b53785a76cffa8928e4d62
cc9/linearmodels-4.16-cp37-cp37m-win_amd64.whl (1.6MB)
Requirement already satisfied: scipy in c:\users\sourav ch\anaconda3\lib\site-packages (from
lmdiag) (1.3.1)
Requirement already satisfied: numpy in c:\users\sourav ch\anaconda3\lib\site-packages (from
lmdiag) (1.16.5)
Requirement already satisfied: statsmodels in c:\users\sourav ch\anaconda3\lib\site-packages (from
lmdiag) (0.10.1)
Requirement already satisfied: matplotlib in c:\users\sourav ch\anaconda3\lib\site-packages (from
lmdiag) (3.1.1)
Requirement already satisfied: pandas in c:\users\sourav ch\anaconda3\lib\site-packages (from
lmdiag) (0.25.1)
Collecting property-cached>=1.6.3 (from linearmodels->lmdiag)
  Downloading
https://files.pythonhosted.org/packages/c0/11/6e91ff5fe0476492f023cebad434a1a34fc513cfa98ddb1f3e5c8
d99/property_cached-1.6.3-py2.py3-none-any.whl
Collecting mypy-extensions>=0.4 (from linearmodels->lmdiag)
  Downloading
https://files.pythonhosted.org/packages/5c/eb/975c7c080f3223a5cdaff09612f3a5221e4ba534f7039db34c35c
6a5/mypy_extensions-0.4.3-py2.py3-none-any.whl
Collecting Cython>=0.29.14 (from linearmodels->lmdiag)
  Downloading
https://files.pythonhosted.org/packages/1f/be/b14be5c3ad1ff73096b518be1538282f053ec34faaca60a8753d9
e93/Cython-0.29.14-cp37-cp37m-win_amd64.whl (1.7MB)
Requirement already satisfied: patsy in c:\users\sourav ch\anaconda3\lib\site-packages (from
linearmodels->lmdiag) (0.5.1)
Requirement already satisfied: cycler>=0.10 in c:\users\sourav ch\anaconda3\lib\site-packages
(from matplotlib->lmdiag) (0.10.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\sourav ch\anaconda3\lib\site-packages
(from matplotlib->lmdiag) (1.1.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in c:\users\sourav
ch\anaconda3\lib\site-packages (from matplotlib->lmdiag) (2.4.2)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\sourav ch\anaconda3\lib\site-
packages (from matplotlib->lmdiag) (2.8.0)
Requirement already satisfied: pytz>=2017.2 in c:\users\sourav ch\anaconda3\lib\site-packages
(from pandas->lmdiag) (2019.3)
Requirement already satisfied: six in c:\users\sourav ch\anaconda3\lib\site-packages (from patsy->
linearmodels->lmdiag) (1.12.0)
Requirement already satisfied: setuptools in c:\users\sourav ch\anaconda3\lib\site-packages (from
kiwisolver>=1.0.1->matplotlib->lmdiag) (41.4.0)
Installing collected packages: property-cached, mypy-extensions, Cython, linearmodels, lmdiag
  Found existing installation: Cython 0.29.13
    Uninstalling Cython-0.29.13:
      Successfully uninstalled Cython-0.29.13
Successfully installed Cython-0.29.14 linearmodels-4.16 lmdiag-0.3.7 mypy-extensions-0.4.3
property-cached-1.6.3
```

In [59]:

```
import lmdiag
import matplotlib.pyplot as plt
lmdiag.plot(model)
plt.show()
```





- **from the above diag plot we can say that the data follows a near normal distribution and we can try to fit linear model .**
- **It also gives us the information that there is no "Heteroscadascticity".**
- **Residuals vs the fitted also do not show any diff patterns so theyy almost follow a common trend .**

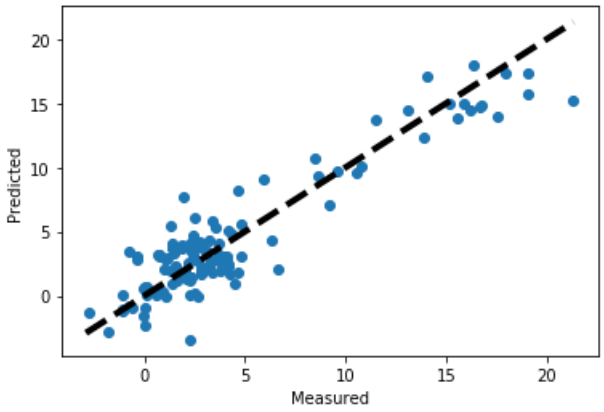
In [82]:

```
####model 1: using scikit learn to predict the MSe metric for model evaluation



X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size = 0.3, random_state = 10)
reg = LinearRegression()
m1 = reg.fit(X_train, Y_train)
#Predict the model on train
pred_train = m1.predict(X_train)
print("MSE on train: ", mean_squared_error(Y_train,pred_train))
print("RSquared: ",m1.score(X_train,Y_train))
#Predict on test
pred_test = m1.predict(X_test)
print("MSE on test: ", mean_squared_error(Y_test,pred_test))
print("RSquared: ",m1.score(X_test,Y_test))
```

```
MSE on train:  2.7102603797808373
RSquared:  0.9048023720966871
MSE on test:  3.9235577435999858
RSquared:  0.8552997473777733
```

In [264]:

```
fig, ax = plt.subplots()
ax.scatter(Y_test, pred_test)
ax.plot([Y.min(), Y.max()], [Y.min(), Y.max()], 'k--', lw=4)
ax.set_xlabel('Measured')
ax.set_ylabel('Predicted')
plt.show()
```



In [24]:

```
df4 = df1.copy()
```

In [23]:

```
#df4.drop(columns = {'subregion',"Nr"}, inplace = True) # removing these columns increased mse so
i did not run this code
```

model2 : using p-value backward elimination we have the

following model and scores

In [25]:

```
x = df4.drop("vot19_14", axis = 1)
Y = df4["vot19_14"]




cols = list(x.columns)
pmax = 1
while (len(cols)>0):
    p= []
    X_1 = x[cols]
    X_1 = sm.add_constant(X_1)
    model2 = sm.OLS(Y,X_1).fit()
    p = pd.Series(model2.pvalues.values[1:],index = cols)
    pmax = max(p)
    print('pmax:',pmax)
    feature_with_p_max = p.idxmax()
    print('feature_pmax',feature_with_p_max)
    if(pmax>0.05):
        cols.remove(feature_with_p_max)

    else:
        break
selected_features_BE = cols
print(selected_features_BE)
```

```
C:\Users\SOURAV CH\Anaconda3\lib\site-packages\numpy\core\fromnumeric.py:2389: FutureWarning: Meth
od .ptp is deprecated and will be removed in a future version. Use numpy.ptp instead.
  return ptp(axis=axis, out=out, **kwargs)
```

```
pmsx: 0.9823014416785996
feature_pmax insolvencies_2012
pmsx: 0.9751402790586139
feature_pmax insolvencies_per_1000_2012
pmsx: 0.976312864213492
feature_pmax business_reg_2017_2012
pmsx: 0.974295537685532
feature_pmax debt_2014
pmsx: 0.9679103543476841
feature_pmax turnout19
pmsx: 0.9470156728153247
feature_pmax age_25_34_2012
pmsx: 0.9457399426133217
feature_pmax empl_total_2018_2012
pmsx: 0.9479636414566406
feature_pmax empl_manuf_2018_2012
pmsx: 0.949584695398314
feature_pmax foreign_suspects_2016
pmsx: 0.9433821248657293
feature_pmax debt_2018
pmsx: 0.9343899013975979
feature_pmax ag4_60_74_2017
pmsx: 0.9293552031382638
feature_pmax debt_2015
pmsx: 0.9058374210164626
feature_pmax empl_oth_service_2018
pmsx: 0.8878620367978666
feature_pmax disposable_inc_2016
pmsx: 0.8667767907475088
feature_pmax foreigners_2012
pmsx: 0.8619620449332492
feature_pmax empl_agr_2012
pmsx: 0.8423788710464833
feature_pmax business_reg_2012
pmsx: 0.9031971761852674
feature_pmax business_delist_2012
pmsx: 0.8538770117932553
feature_pmax debt_2017
pmsx: 0.9300477480470293
feature_pmax debt_2016
pmsx: 0.8287923441805888
```

```
feature_pmax net_migration_2017_2012
pmsx: 0.832103871485715
feature_pmax unempl_female_2013
pmsx: 0.8129612471043786
feature_pmax hartz_total_2013
pmsx: 0.7984813668959626
feature_pmax net_migration_2012
pmsx: 0.7904412428985786
feature_pmax total_suspects_2015
pmsx: 0.8659751055058942
feature_pmax f_crime_2015
pmsx: 0.7689382985107848
feature_pmax foreign_suspects_2018
pmsx: 0.8088518948240919
feature_pmax total_suspects_2018
pmsx: 0.7844737903753496
feature_pmax unempl_15_19_2019
pmsx: 0.7769804638934206
feature_pmax foreigners_2017_2012
pmsx: 0.6987175442239892
feature_pmax birth_balance_2017
pmsx: 0.7385590891284994
feature_pmax graduates_sec_2012
pmsx: 0.7263427024913935
feature_pmax foreigners_2017
pmsx: 0.6635296388411157
feature_pmax population_density_2017_2012
pmsx: 0.655230343958111
feature_pmax dwellings_new_2017
pmsx: 0.6845138717746047
feature_pmax empl_soc_sec_total_2012
pmsx: 0.8019538096050057
feature_pmax empl_total_2018
pmsx: 0.6316625253034049
feature_pmax age_35_59_2017_2012
pmsx: 0.6286762818896834
feature_pmax age_25_34_2017_2012
pmsx: 0.6094106552727401
feature_pmax total_suspects_2013
pmsx: 0.5875363663101183
feature_pmax area_2012
pmsx: 0.5394363139836856
feature_pmax f_crime_2016
pmsx: 0.5546190522982204
feature_pmax population_2012
pmsx: 0.46283500926569354
feature_pmax hartz_no_empl_2018_2013
pmsx: 0.5036827937753559
feature_pmax hartz_no_empl_2018
pmsx: 0.5174393015822581
feature_pmax f_crime_2014
pmsx: 0.47976281733596504
feature_pmax protection_rejected_2017
pmsx: 0.911631503572509
feature_pmax protection_accepted_2017
pmsx: 0.6240419233086596
feature_pmax protection_open_2017
pmsx: 0.4673131003419756
feature_pmax foreign_suspects_2014
pmsx: 0.5198951574380 91
feature_pmax total_suspects_2014
pmsx: 0.4116907700877702
feature_pmax germans_2017
pmsx: 0.38064458104092536
feature_pmax graduates_without_secondary_2017
pmsx: 0.3861572790524159
feature_pmax age_to_18_2017_2012
pmsx: 0.46152665683809024
feature_pmax dwellings_2017_2012
pmsx: 0.43650639560635274
feature_pmax foreign_suspects_2013
pmsx: 0.3584931891094483
feature_pmax trade_tax_per_inh_2012
pmsx: 0.6419766984400108
feature_pmax gdp_2016
pmsx: 0.3610552871237611
feature pmax graduates voc 2017
```

pmsx: 0.38646444911569056
feature_pmax total_suspects_2017
pmsx: 0.3460927918986908
feature_pmax Absolventen/Abgänger allgemeinbildender Schulen 2017 - insgesamt ohne Externe (je 100
0 Einwohner)
pmsx: 0.32595615876885786
feature_pmax empl_service_2018
pmsx: 0.3076091118046995
feature_pmax age_75_more_2017
pmsx: 0.35783118817438975
feature_pmax ag4_60_74_2017_2012
pmsx: 0.6083802506967217
feature_pmax age_75_more_2017_2012
pmsx: 0.32027668018626576
feature_pmax age_18_24_2017
pmsx: 0.32547877993967367
feature_pmax child_day_care_2018
pmsx: 0.2643490244163826
feature_pmax population_2017
pmsx: 0.30034009431974984
feature_pmax male_2012
pmsx: 0.32031601855089153
feature_pmax empl_oth_service_2018_2012
pmsx: 0.24852927609120157
feature_pmax insolvencies_2017_2012
pmsx: 0.20365757132516063
feature_pmax protection_total_2017
pmsx: 0.20350543200290558
feature_pmax hartz_no_empl_2013
pmsx: 0.1605249840378828
feature_pmax Nr
pmsx: 0.1612982170878141
feature_pmax area_2017
pmsx: 0.2068166493381489
feature_pmax total_suspects_2012
pmsx: 0.31050631310597515
feature_pmax total_suspects_2016
pmsx: 0.1618737998341341
feature_pmax hartz_total_2018_2013
pmsx: 0.17108247200437782
feature_pmax graduates_secondary_2017
pmsx: 0.14877894554023494
feature_pmax f_crime_2018
pmsx: 0.15219606587059153
feature_pmax debt_2013
pmsx: 0.21607805282614473
feature_pmax turnout19_14
pmsx: 0.16132379008501704
feature_pmax f_crime_2012
pmsx: 0.28254481441675827
feature_pmax foreign_suspects_2017
pmsx: 0.25244928781323106
feature_pmax f_crime_2017
pmsx: 0.09818221600324856
feature_pmax birth_balance_2012
pmsx: 0.09558090120971266
feature_pmax age_to_18_2017
pmsx: 0.08630062342599532
feature_pmax dwellings_new_2012
pmsx: 0.08298203247457205
feature_pmax dwellings_2012
pmsx: 0.13747168662214862
feature_pmax population_density_2017
pmsx: 0.08852110328133143
feature_pmax graduates_without_secondary_2017_2012
pmsx: 0.14146611114200128
feature_pmax graduates_without_secondary_2012
pmsx: 0.20589961777903096
feature_pmax graduates_lower_secondary_2012
pmsx: 0.37783060194132645
feature_pmax graduates_uni_2012
pmsx: 0.33298062177223486
feature_pmax graduates_secondary_2012
pmsx: 0.0761407978674134
feature_pmax empl_service_2018_2012
pmsx: 0.06689255705138185
feature_pmax net_migration_2017

```
pmsx: 0.051830349310663014
feature_pmax empl_com_hotel_2018_2012
pmsx: 0.05027895551829242
feature_pmax empl_manuf_2018
pmsx: 0.058006467980172796
feature_pmax empl_agr_2018_2012
pmsx: 0.06787483256175696
feature_pmax vehicles_2018_2013
pmsx: 0.12956352777424276
feature_pmax age_18_24_2017_2012
pmsx: 0.059701709898329826
feature_pmax unempl_male_2019_2013
pmsx: 0.02702105685236611
feature_pmax unempl_female_2019_2013
['subregion', 'turnout14', 'ove18_13', 'age_25_34_2017', 'age_35_59_2017', 'dwellings_2017',
'space_per_app_2017', 'space_per_inh_2017', 'vehicles_2018', 'graduates_lower_secondary_2017', 'gr
aduates_higher_2017', 'business_reg_2017', 'insolvencies_2017', 'empl_agr_2018',
'empl_com_hotel_2018', 'hartz_total_2018', 'hartz_foreign_2018', 'unempl_total_2019',
'unempl_male_2019', 'unempl_female_2019', 'unempl_55_64_2019', 'birth_balance_2017_2012',
'graduates_lower_secondary_2017_2012', 'graduates_secondary_2017_2012', 'dwellings_new_2017_2012',
'unempl_total_2019_2013', 'unempl_female_2019_2013', 'population_density_2012', 'age_to_18_2012',
'age_18_24_2012', 'age_35_59_2012', 'ag4_60_74_2012', 'age_75_more_2012', 'vehicles_2013',
'mining_manuf_2012', 'empl_manuf_2012', 'empl_com_hotel_2012', 'empl_service_2012',
'empl_oth_service_2012', 'unempl_total_2013', 'unempl_male_2013', 'foreign_suspects_2015',
'f_crime_2013', 'foreign_suspects_2012']
```

In [322]:

```
p
```

Out[322]:

```
subregion                                5.345279e-06
turnout14                                6.577139e-10
ove18_13                                 3.402800e-06
age_25_34_2017                           1.372517e-05
age_35_59_2017                           7.938381e-04
dwellings_2017                           4.763489e-06
space_per_app_2017                       1.826875e-06
space_per_inh_2017                       6.895299e-08
vehicles_2018                            3.444644e-03
graduates_lower_secondary_2017           1.817984e-06
graduates_higher_2017                    4.716193e-04
business_reg_2017                        1.875808e-08
insolvencies_2017                        2.348688e-04
empl_agr_2018                            1.072381e-03
empl_com_hotel_2018                      1.318474e-03
hartz_total_2018                         3.551546e-04
hartz_foreign_2018                       9.724971e-05
unempl_total_2019                        1.148729e-02
unempl_male_2019                         2.424854e-02
unempl_female_2019                       2.574754e-03
unempl_55_64_2019                        6.821649e-04
birth_balance_2017_2012                  7.183837e-03
graduates_lower_secondary_2017_2012      2.471338e-03
graduates_secondary_2017_2012            7.166972e-03
dwellings_new_2017_2012                  8.130134e-03
unempl_total_2019_2013                   2.644346e-02
unempl_female_2019_2013                  2.702106e-02
population_density_2012                  1.607579e-02
age_to_18_2012                           5.210391e-10
age_18_24_2012                           2.229763e-14
age_35_59_2012                           1.071903e-02
ag4_60_74_2012                           1.092947e-09
age_75_more_2012                         8.612342e-06
vehicles_2013                            7.055586e-04
mining_manuf_2012                        1.396521e-02
empl_manuf_2012                          7.550964e-03
empl_com_hotel_2012                      1.071291e-03
empl_service_2012                        5.676659e-03
empl_oth_service_2012                    5.092879e-03
unempl_total_2013                        3.924841e-05
unempl_male_2013                         6.459955e-05
foreign_suspects_2015                    1.297292e-02
f_crime_2013                             4.492678e-03
foreign_suspects_2012                    1.738121e-02
```

foreign_suspects_2012 1.756121e-02
dtype: float64

In [26]:

```
len(selected_features_BE)
```

Out[26]:

44

In [27]:

```
model2.summary()
```

Out[27]:

OLS Regression Results

Dep. Variable:	vot19_14	**R-squared:**	0.933
Model:	OLS	**Adj. R-squared:**	0.925
Method:	Least Squares	**F-statistic:**	113.5
Date:	Wed, 29 Jan 2020	**Prob (F-statistic):**	9.95e-183
Time:	12:54:55	**Log-Likelihood:**	-694.26
No. Observations:	401	**AIC:**	1479.
Df Residuals:	356	**BIC:**	1658.
Df Model:	44		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	179.2441	32.173	5.571	0.000	115.971	242.518
subregion	0.2111	0.046	4.621	0.000	0.121	0.301
turnout14	9.8914	1.558	6.350	0.000	6.828	12.955
ove18_13	0.8871	0.188	4.720	0.000	0.517	1.257
age_25_34_2017	-0.9815	0.223	-4.410	0.000	-1.419	-0.544
age_35_59_2017	-0.7039	0.208	-3.384	0.001	-1.113	-0.295
dwellings_2017	0.0649	0.014	4.646	0.000	0.037	0.092
space_per_app_2017	0.3533	0.073	4.853	0.000	0.210	0.497
space_per_inh_2017	-0.8240	0.150	-5.510	0.000	-1.118	-0.530
vehicles_2018	-0.0188	0.006	-2.945	0.003	-0.031	-0.006
graduates_lower_secondary_2017	-0.1260	0.026	-4.854	0.000	-0.177	-0.075
graduates_higher_2017	-0.0592	0.017	-3.529	0.000	-0.092	-0.026
business_reg_2017	-0.1388	0.024	-5.754	0.000	-0.186	-0.091
insolvencies_2017	0.4166	0.112	3.716	0.000	0.196	0.637
empl_agr_2018	-0.9002	0.273	-3.298	0.001	-1.437	-0.363
empl_com_hotel_2018	0.1774	0.055	3.238	0.001	0.070	0.285
hartz_total_2018	0.0391	0.011	3.606	0.000	0.018	0.060
hartz_foreign_2018	-0.0671	0.017	-3.942	0.000	-0.101	-0.034
unempl_total_2019	5.5426	2.182	2.541	0.011	1.252	9.833
unempl_male_2019	-2.6297	1.162	-2.263	0.024	-4.915	-0.344
unempl_female_2019	-3.1808	1.048	-3.036	0.003	-5.241	-1.120
unempl_55_64_2019	0.5143	0.150	3.427	0.001	0.219	0.809
birth_balance_2017_2012	0.0006	0.000	2.704	0.007	0.000	0.001
graduates_lower_secondary_2017_2012	0.0099	0.003	3.049	0.002	0.004	0.016
graduates_secondary_2017_2012	-0.0127	0.005	-2.705	0.007	-0.022	-0.003
dwellings_new_2017_2012	0.0013	0.000	2.662	0.008	0.000	0.002

unempl_total_2019_2013	-0.0100	0.005	-2.229	0.026	-0.019	-0.001
unempl_female_2019_2013	0.0235	0.011	2.220	0.027	0.003	0.044
population_density_2012	-0.0007	0.000	-2.419	0.016	-0.001	-0.000
age_to_18_2012	-1.4904	0.233	-6.390	0.000	-1.949	-1.032
age_18_24_2012	-2.0951	0.263	-7.966	0.000	-2.612	-1.578
age_35_59_2012	-0.5868	0.229	-2.565	0.011	-1.037	-0.137
ag4_60_74_2012	-1.5377	0.246	-6.262	0.000	-2.021	-1.055
age_75_more_2012	-1.0773	0.239	-4.515	0.000	-1.547	-0.608
vehicles_2013	0.0243	0.007	3.417	0.001	0.010	0.038
mining_manuf_2012	1.1095	0.449	2.470	0.014	0.226	1.993
empl_manuf_2012	-0.6631	0.247	-2.687	0.008	-1.148	-0.178
empl_com_hotel_2012	-0.8386	0.254	-3.298	0.001	-1.339	-0.339
empl_service_2012	-0.6869	0.247	-2.783	0.006	-1.172	-0.201
empl_oth_service_2012	-0.7031	0.249	-2.819	0.005	-1.194	-0.213
unempl_total_2013	-1.4867	0.357	-4.164	0.000	-2.189	-0.785
unempl_male_2013	1.3974	0.346	4.043	0.000	0.718	2.077
foreign_suspects_2015	0.0046	0.002	2.497	0.013	0.001	0.008
f_crime_2013	-0.0336	0.012	-2.860	0.004	-0.057	-0.010
foreign_suspects_2012	-0.0053	0.002	-2.390	0.017	-0.010	-0.001

Omnibus:	7.274	**Durbin-Watson:**	1.837
Prob(Omnibus):	0.026	**Jarque-Bera (JB):**	9.411
Skew:	0.160	**Prob(JB):**	0.00905
Kurtosis:	3.679	**Cond. No.**	5.75e+05

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 5.75e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

In [30]:

```
df_4 = df4[selected_features_BE]
```

In [31]:

```
X = df_4
Y = df4["vot19_14"]
x1 = sm.add_constant(X) #Adding the constant
model_2 = sm.OLS(Y,x1).fit() # fitting the model
print(model_2.summary()) # model summary
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                vot19_14   R-squared:                       0.933
Model:                            OLS   Adj. R-squared:                  0.925
Method:                 Least Squares   F-statistic:                     113.5
Date:                Wed, 29 Jan 2020   Prob (F-statistic):          9.95e-183
Time:                        12:57:07   Log-Likelihood:                -694.26
No. Observations:                 401   AIC:                             1479.
Df Residuals:                     356   BIC:                             1658.
Df Model:                          44
Covariance Type:            nonrobust
==============================================================================================================

                                       coef    std err          t      P>|t|      [0.025
```

```
0.975]
------------------------------------------------------------------------------------------------
const                                179.2441     32.173      5.571      0.000     115.971     242
18
subregion                              0.2111      0.046      4.621      0.000       0.121       0
01
turnout14                              9.8914      1.558      6.350      0.000       6.828      12
55
ove18_13                               0.8871      0.188      4.720      0.000       0.517       1
57
age_25_34_2017                        -0.9815      0.223     -4.410      0.000      -1.419      -0
44
age_35_59_2017                        -0.7039      0.208     -3.384      0.001      -1.113      -0
95
dwellings_2017                         0.0649      0.014      4.646      0.000       0.037       0
92
space_per_app_2017                     0.3533      0.073      4.853      0.000       0.210       0
97
space_per_inh_2017                    -0.8240      0.150     -5.510      0.000      -1.118      -0
530
vehicles_2018                         -0.0188      0.006     -2.945      0.003      -0.031      -0
06
graduates_lower_secondary_2017        -0.1260      0.026     -4.854      0.000      -0.177      -0
075
graduates_higher_2017                 -0.0592      0.017     -3.529      0.000      -0.092      -0
026
business_reg_2017                     -0.1388      0.024     -5.754      0.000      -0.186      -0
091
insolvencies_2017                      0.4166      0.112      3.716      0.000       0.196       0
37
empl_agr_2018                         -0.9002      0.273     -3.298      0.001      -1.437      -0
63
empl_com_hotel_2018                    0.1774      0.055      3.238      0.001       0.070       0
85
hartz_total_2018                       0.0391      0.011      3.606      0.000       0.018       0
60
hartz_foreign_2018                    -0.0671      0.017     -3.942      0.000      -0.101      -0
034
unempl_total_2019                      5.5426      2.182      2.541      0.011       1.252       9
33
unempl_male_2019                      -2.6297      1.162     -2.263      0.024      -4.915      -
0.344
unempl_female_2019                    -3.1808      1.048     -3.036      0.003      -5.241      -1
120
unempl_55_64_2019                      0.5143      0.150      3.427      0.001       0.219       0
09
birth_balance_2017_2012                0.0006      0.000      2.704      0.007       0.000
0.001
graduates_lower_secondary_2017_2012    0.0099      0.003      3.049      0.002       0.004       0
016
graduates_secondary_2017_2012         -0.0127      0.005     -2.705      0.007      -0.022      -0
003
dwellings_new_2017_2012                0.0013      0.000      2.662      0.008       0.000
0.002
unempl_total_2019_2013                -0.0100      0.005     -2.229      0.026      -0.019      -0
001
unempl_female_2019_2013                0.0235      0.011      2.220      0.027       0.003
0.044
population_density_2012               -0.0007      0.000     -2.419      0.016      -0.001      -0
000
age_to_18_2012                        -1.4904      0.233     -6.390      0.000      -1.949      -1
32
age_18_24_2012                        -2.0951      0.263     -7.966      0.000      -2.612      -1
78
age_35_59_2012                        -0.5868      0.229     -2.565      0.011      -1.037      -0
37
ag4_60_74_2012                        -1.5377      0.246     -6.262      0.000      -2.021      -1
55
age_75_more_2012                      -1.0773      0.239     -4.515      0.000      -1.547      -
0.608
vehicles_2013                          0.0243      0.007      3.417      0.001       0.010       0
38
mining_manuf_2012                      1.1095      0.449      2.470      0.014       0.226       1
93
empl_manuf_2012                       -0.6631      0.247     -2.687      0.008      -1.148      -
0.178
empl_com_hotel_2012                   -0.8386      0.254     -3.298      0.001      -1.339      -0
```

```
339
empl_service_2012                    -0.6869      0.247     -2.783      0.006      -1.172      -0
201
empl_oth_service_2012                -0.7031      0.249     -2.819      0.005      -1.194      -0
213
unempl_total_2013                    -1.4867      0.357     -4.164      0.000      -2.189      -0
785
unempl_male_2013                      1.3974      0.346      4.043      0.000       0.718       2
77
foreign_suspects_2015                 0.0046      0.002      2.497      0.013       0.001
0.008
f_crime_2013                         -0.0336      0.012     -2.860      0.004      -0.057      -0
10
foreign_suspects_2012                -0.0053      0.002     -2.390      0.017      -0.010      -0
001
==============================================================================
Omnibus:                        7.274   Durbin-Watson:                   1.837
Prob(Omnibus):                  0.026   Jarque-Bera (JB):                9.411
Skew:                           0.160   Prob(JB):                      0.00905
Kurtosis:                       3.679   Cond. No.                     5.75e+05
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 5.75e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
```

In [335]:

```
vif = pd.DataFrame()
vif["VIF Factor"] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
vif["features"] = X.columns
```

In [32]:

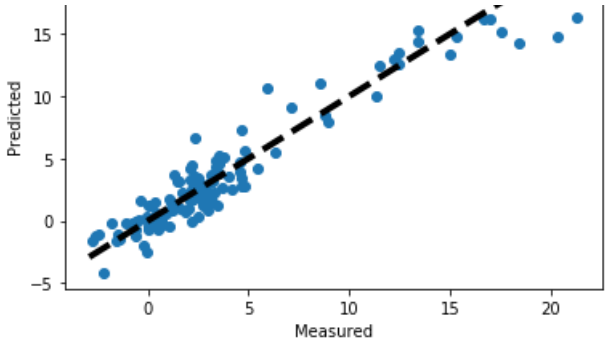
```
# model2 : with p values < 0.05 Backward elimination with  subregion and Nr

X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size = 0.3, random_state = 12)
reg = LinearRegression()
m_2 = reg.fit(X_train, Y_train)
#Predict the model on train
pred_train = m_2.predict(X_train)
print("MSE on train: ", mean_squared_error(Y_train,pred_train))
print("RSquared: ",m_2.score(X_train,Y_train))
#Predict on test
pred_test = m_2.predict(X_test)
print("MSE on test: ", mean_squared_error(Y_test,pred_test))
print("RSquared: ",m_2.score(X_test,Y_test))
```

```
MSE on train:  1.8526466142349656
RSquared:  0.935216740673382
MSE on test:  2.3068059689022884
RSquared:  0.9124425045324986
```

- The MSE of the this model is very good and the score are good as well but when your train and test data gets such a good score we suspect a overfit.
- so lets build a data frame to see the predictions on test .
- lets also visualize our best fit line.

In [271]:

```
fig, ax = plt.subplots()
ax.scatter(Y_test, pred_test)
ax.plot([Y.min(), Y.max()], [Y.min(), Y.max()], 'k--', lw=4)
ax.set_xlabel('Measured')
ax.set_ylabel('Predicted')
plt.show()
```



In [38]:

```
values = pd.DataFrame(pred_test,Y_test)
values
```

Out[38]:

vot19_14	0
2.400414	1.944227
18.334998	18.200895
-0.724483	-0.171693
0.060316	0.004437
-0.329070	1.663082
...	...
2.524834	3.417599
3.027529	1.544795
2.114120	1.584297
2.299495	2.484344
0.252615	0.247788

121 rows × 1 columns

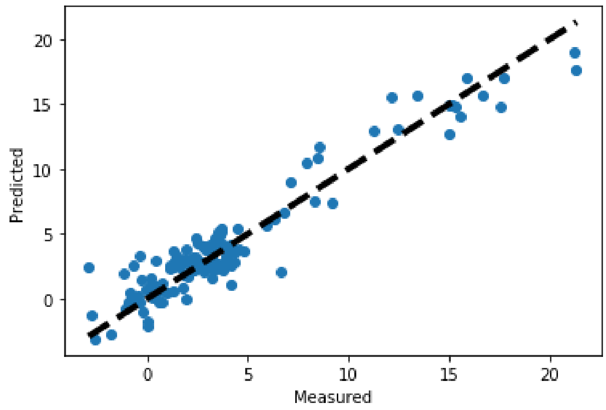
In [272]:

```
# model2 with backward elimination with out subregion and Nr
X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size = 0.3, random_state = 2)
reg = LinearRegression()
m2 = reg.fit(X_train, Y_train)
#Predict the model on train
pred_train = m2.predict(X_train)
print("MSE on train: ", mean_squared_error(Y_train,pred_train))
print("RSquared: ",m2.score(X_train,Y_train))
#Predict on test
pred_test = m2.predict(X_test)
print("MSE on test: ", mean_squared_error(Y_test,pred_test))
print("RSquared: ",m2.score(X_test,Y_test))
```

```
MSE on train:  1.7747006616235133
RSquared:  0.9386584521015146
MSE on test:  2.4192473637960967
RSquared:  0.9062132804166938
```

In [273]:

```
fig, ax = plt.subplots()
ax.scatter(Y_test, pred_test)
ax.plot([Y.min(), Y.max()], [Y.min(), Y.max()], 'k--', lw=4)
ax.set_xlabel('Measured')
ax.set_ylabel('Predicted')
plt.show()
```



model3: Using recursive feature extraction

In [10]:

```
df5 = df1.copy()
```

In [11]:

```
X= df5.drop("vot19_14", axis = 1)
y = df5["vot19_14"]
model3 = LinearRegression()
#Initializing RFE model
rfe = RFE(model3, 150)
#Transforming data using RFE
X_rfe = rfe.fit_transform(X,y)
#Fitting the data to model
model3.fit(X_rfe,y)
print(rfe.support_)
print(rfe.ranking_)
```

```
[ True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True]
[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]
```

Basically the model which has been fit , using rfe is giving rankings based on the feature importance and gives us the best features that actually explain the target.

In [13]:

```
#no of features
nof_list=np.arange(1,150)
high_score=0
#Variable to store the optimum features
nof=0
score_list =[]
for n in range(len(nof_list)):
    X_train, X_test, y_train, y_test = train_test_split(X,y, test_size = 0.3, random_state = 0)
    model3 = LinearRegression()
    rfe = RFE(model3,nof_list[n])
    X_train_rfe = rfe.fit_transform(X_train,y_train)
```

```
    X_test_rfe = rfe.transform(X_test)
    model3.fit(X_train_rfe,y_train)
    score = model3.score(X_test_rfe,y_test)
    score_list.append(score)
    if(score>high_score):
        high_score = score
        nof = nof_list[n]
print("Optimum number of features: %d" %nof)
print("Score with %d features: %f" % (nof, high_score))
```

```
Optimum number of features: 50
Score with 50 features: 0.882529
```

In [14]:

```
cols = list(X.columns)
model3 = LinearRegression()
#Initializing RFE model
rfe = RFE(model3, 50)
#Transforming data using RFE
X_rfe = rfe.fit_transform(X,y)
#Fitting the data to model
model3.fit(X_rfe,y)
temp = pd.Series(rfe.support_,index = cols)
selected_features_rfe = temp[temp==True].index
print(selected_features_rfe)
```

```
Index(['subregion', 'turnout14', 'turnout19_14', 'debt_2013', 'debt_2014',
       'debt_2017', 'debt_2018', 'ove18_13', 'foreigners_2017',
       'birth_balance_2017', 'age_to_18_2017', 'age_18_24_2017',
       'age_25_34_2017', 'age_35_59_2017', 'ag4_60_74_2017',
       'age_75_more_2017', 'protection_open_2017', 'protection_accepted_2017',
       'protection_rejected_2017', 'dwellings_new_2017',
       'graduates_without_secondary_2017', 'graduates_lower_secondary_2017',
       'graduates_secondary_2017', 'graduates_higher_2017',
       'business_reg_2017', 'insolvencies_2017', 'empl_agr_2018',
       'empl_com_hotel_2018', 'unempl_total_2019', 'unempl_male_2019',
       'unempl_female_2019', 'unempl_55_64_2019', 'birth_balance_2012',
       'age_to_18_2012', 'age_18_24_2012', 'age_35_59_2012', 'ag4_60_74_2012',
       'age_75_more_2012', 'graduates_without_secondary_2012',
       'graduates_lower_secondary_2012', 'graduates_secondary_2012',
       'graduates_uni_2012', 'mining_manuf_2012', 'insolvencies_2012',
       'empl_manuf_2012', 'empl_com_hotel_2012', 'empl_service_2012',
       'empl_oth_service_2012', 'unempl_total_2013', 'unempl_male_2013'],
      dtype='object')
```

In [15]:

```
df_5 = df5[selected_features_rfe]
```

In [213]:

```

```

Out[213]:

	subregion	turnout14	turnout19_14	debt_2013	debt_2014	debt_2017	debt_2018	ove18_13	foreigners_2017	birth_balance_2017
0	1.0	0.357400	0.205520	16.41	16.40	16.21	16.24	-0.17	13.3	-1.7
1	1.0	0.402589	0.186015	12.04	12.03	12.16	11.96	-0.08	11.3	0.1
2	1.0	0.376398	0.169726	15.25	15.59	15.04	14.76	-0.49	9.7	-3.6
3	1.0	0.453649	0.028556	16.61	16.94	17.80	18.09	1.48	11.0	-5.4
4	1.0	0.397193	0.146915	12.52	12.80	12.78	12.84	0.32	5.6	-5.3
5	1.0	0.463842	0.138119	10.16	10.29	10.17	10.23	0.07	7.5	-3.0
6	1.0	0.411905	0.176788	10.09	10.32	9.82	9.93	-0.16	6.8	-3.6
7	1.0	0.424788	0.159476	10.63	10.80	10.62	10.47	-0.16	5.5	-6.9
8	1.0	0.457528	0.172180	9.67	9.64	9.60	9.62	-0.05	10.1	-2.2

	subregion	unemployment_2017	turnout_2017	debt_2013	debt_2014	debt_2015	debt_2016	over18_17	foreigners_2017	birth_balance_2017
9	1.0	0.469134	0.169734	9.63	9.68	9.77	9.75	0.17	4.1	-3.7
10	1.0	0.459591	0.165319	9.16	9.17	9.41	9.38	0.22	5.0	-3.6
11	1.0	0.421130	0.177120	10.45	10.67	10.64	10.61	0.16	5.1	-3.8
12	1.0	0.423453	0.171872	10.11	10.08	9.86	9.82	-0.29	8.1	-2.2
13	1.0	0.420752	0.156371	11.43	11.66	12.18	12.15	0.72	6.6	-3.7
14	1.0	0.503786	0.150481	7.99	7.93	7.63	7.55	-0.44	7.0	-3.1
15	2.0	0.435025	0.183690	10.92	10.81	10.61	10.62	-0.30	16.2	1.9
16	3.0	0.512913	0.128824	10.62	10.47	9.61	9.67	-0.95	9.7	-2.1
17	3.0	0.459825	0.051813	12.16	12.51	13.24	13.49	1.33	16.9	-3.3
18	3.0	0.412644	0.152323	8.17	7.99	7.69	7.74	-0.43	14.3	-1.7
19	3.0	0.477766	0.129914	9.02	8.98	8.85	8.74	-0.28	6.2	-0.8
20	3.0	0.383769	0.200491	12.74	12.88	12.47	12.51	-0.23	8.9	-8.5
21	3.0	0.423608	0.160607	10.97	11.19	11.04	11.21	0.24	6.6	-5.5
22	3.0	0.481927	0.079024	10.80	10.98	10.66	10.68	-0.12	5.9	-6.7
23	3.0	0.468641	0.150865	9.78	10.00	10.33	10.41	0.63	7.2	-2.9
24	3.0	0.525676	0.118711	9.13	9.22	9.10	9.09	-0.04	5.9	-4.4
25	3.0	0.502767	0.099959	9.82	9.73	9.58	9.57	-0.25	8.1	-3.6
26	3.0	0.486508	0.153106	12.01	11.93	11.50	11.45	-0.56	13.1	-1.2
27	3.0	0.495989	0.125992	9.23	9.28	9.41	9.39	0.16	7.8	-2.9
28	3.0	0.491912	0.092370	12.23	12.40	12.78	12.93	0.70	9.7	-5.6
29	3.0	0.484514	0.139769	10.98	11.07	10.80	10.65	-0.33	7.6	-5.7
...	...	...	...	...	...	...	...	...	...	...
371	15.0	0.407577	0.127519	11.99	12.42	12.39	12.34	0.35	3.0	-8.4
372	15.0	0.479496	0.068303	11.74	11.82	11.88	11.94	0.20	3.4	-6.6
373	15.0	0.452224	0.084795	11.27	11.62	12.22	12.16	0.89	2.9	-9.7
374	15.0	0.447176	0.123932	11.30	11.45	11.60	11.59	0.29	3.5	-5.9
375	15.0	0.405973	0.094412	12.33	12.54	12.95	13.10	0.77	3.1	-9.6
376	15.0	0.427994	0.115690	11.92	12.15	12.42	12.33	0.41	3.7	-6.6
377	15.0	0.467126	0.108452	9.96	10.29	10.48	10.56	0.60	3.2	-7.5
378	16.0	0.479950	0.127485	11.48	11.45	11.17	11.19	-0.29	7.4	-1.4
379	16.0	0.435021	0.132615	11.62	11.90	11.77	11.85	0.23	5.8	-6.6
380	16.0	0.522503	0.127272	5.81	5.70	5.94	5.91	0.10	8.9	0.9
381	16.0	0.451281	0.113586	10.03	10.11	10.55	10.81	0.78	7.0	-9.3
382	16.0	0.499509	0.130687	10.79	10.49	10.75	10.74	-0.05	8.2	-2.7
383	16.0	0.516287	0.079205	11.87	12.16	12.06	11.91	0.04	7.9	-5.5
384	16.0	0.573239	0.085326	6.19	6.27	6.32	6.14	-0.05	3.3	-2.5
385	16.0	0.485793	0.102730	9.47	9.82	10.19	10.42	0.95	4.2	-6.4
386	16.0	0.544328	0.081478	7.53	7.68	8.05	8.01	0.48	2.8	-5.4
387	16.0	0.496123	0.098784	9.80	9.95	9.91	10.21	0.41	3.5	-5.7
388	16.0	0.516722	0.077842	10.88	10.97	11.40	11.56	0.68	2.8	-8.6
389	16.0	0.521090	0.090600	8.92	9.11	9.54	9.54	0.62	3.2	-6.3
390	16.0	0.519206	0.079432	9.66	9.73	10.45	10.59	0.93	5.4	-5.5
391	16.0	0.537037	0.074817	9.03	9.21	9.47	9.54	0.51	2.6	-4.5
392	16.0	0.524583	0.121694	8.01	8.25	8.35	8.59	0.58	2.9	-6.5
393	16.0	0.528120	0.106144	9.18	9.21	9.49	9.57	0.39	5.3	-5.2
394	16.0	0.565730	0.092031	9.30	9.36	9.49	9.46	0.16	3.5	-4.1
395	16.0	0.473913	0.100873	8.83	8.81	8.87	8.93	0.10	4.1	-8.5
396	16.0	0.508175	0.101750	7.45	7.81	8.18	8.23	0.78	2.7	-8.5
397	16.0	0.584836	0.075714	7.40	7.40	7.43	7.50	0.10	2.9	-4.5
398	16.0	0.541176	0.093087	7.48	7.56	7.52	7.47	-0.01	3.3	-6.7

	subregion	turnout14	turnout19_14	debt_2013	debt_2014	debt_2017	debt_2018	ove18_13	foreigners_2017	birth_balance_2017
399	16.0	0.553222	0.088169	7.42	7.77	7.80	7.76	0.34	2.3	-7.7
400	16.0	0.483017	0.082643	8.29	8.66	8.87	8.95	0.66	3.2	-7.9

401 rows × 50 columns

In [16]:

```
X = df_5
Y = df5["vot19_14"]
x1 = sm.add_constant(X) #Adding the constant
model3 = sm.OLS(Y,x1).fit() # fitting the model
print(model3.summary()) # model summary
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                vot19_14   R-squared:                       0.924
Model:                            OLS   Adj. R-squared:                  0.913
Method:                 Least Squares   F-statistic:                     86.65
Date:                Tue, 11 Feb 2020   Prob (F-statistic):          2.85e-167
Time:                        12:49:32   Log-Likelihood:                -721.79
No. Observations:                 401   AIC:                             1544.
Df Residuals:                     351   BIC:                             1743.
Df Model:                          49
Covariance Type:            nonrobust
================================================================================================
                                   coef    std err          t      P>|t|      [0.025      0.975
------------------------------------------------------------------------------------------------
const                           26.3911    137.642      0.192      0.848    -244.316    297.09
subregion                        0.2039      0.053      3.865      0.000       0.100      0.30
turnout14                        4.5891      2.901      1.582      0.115      -1.117     10.29
turnout19_14                    -4.3366      4.616     -0.939      0.348     -13.415      4.74
debt_2013                       -0.5717      0.501     -1.142      0.254      -1.556      0.41
debt_2014                        1.0318      0.815      1.267      0.206      -0.570      2.63
debt_2017                       -0.7914      0.795     -0.995      0.320      -2.356      0.77
debt_2018                        0.2807      0.508      0.553      0.581      -0.718      1.27
ove18_13                         0.8524      0.335      2.548      0.011       0.194      1.51
foreigners_2017                 -0.1232      0.042     -2.960      0.003      -0.205     -0.04
birth_balance_2017              -0.2002      0.127     -1.578      0.115      -0.450      0.04
age_to_18_2017                   0.7102      1.163      0.610      0.542      -1.578      2.99
age_18_24_2017                  -1.3520      1.146     -1.179      0.239      -3.607      0.90
age_25_34_2017                  -1.2289      1.145     -1.073      0.284      -3.481      1.02
age_35_59_2017                  -1.4633      1.132     -1.293      0.197      -3.690      0.76
ag4_60_74_2017                  -0.6139      1.115     -0.550      0.582      -2.808      1.58
age_75_more_2017                -0.6423      1.118     -0.574      0.566      -2.841      1.55
protection_open_2017             1.9861      0.750      2.650      0.008       0.512      3.46
protection_accepted_2017         1.9684      0.748      2.630      0.009       0.497      3.44
protection_rejected_2017         1.9895      0.748      2.660      0.008       0.519      3.46
dwellings_new_2017               0.0978      0.067      1.452      0.147      -0.035      0.23
graduates_without_secondary_2017 -0.5100     1.079     -0.473      0.637      -2.633      1.61
graduates_lower_secondary_2017   -0.6096     1.076     -0.567      0.571      -2.726      1.50
graduates_secondary_2017         -0.5832     1.073     -0.543      0.587      -2.694      1.52
graduates_higher_2017            -0.5820     1.073     -0.542      0.588      -2.692      1.52
business_reg_2017                -0.1603     0.025     -6.318      0.000      -0.210     -0.11
insolvencies_2017                 0.5181     0.120      4.319      0.000       0.282      0.75
empl_agr_2018                    -0.5051     0.281     -1.796      0.073      -1.058      0.04
empl_com_hotel_2018               0.1436     0.059      2.422      0.016       0.027      0.26
unempl_total_2019                 2.8269     2.367      1.195      0.233      -1.828      7.48
unempl_male_2019                 -1.4261     1.263     -1.129      0.260      -3.910      1.05
unempl_female_2019               -1.5164     1.138     -1.333      0.184      -3.754      0.72
unempl_55_64_2019                 0.4609     0.158      2.923      0.004       0.151      0.77
birth_balance_2012               -0.3188     0.154     -2.070      0.039      -0.622     -0.01
age_to_18_2012                   -2.1907     0.316     -6.929      0.000      -2.812     -1.56
age_18_24_2012                   -1.2463     0.430     -2.898      0.004      -2.092     -0.40
age_35_59_2012                   -0.5382     0.280     -1.920      0.056      -1.089      0.01
ag4_60_74_2012                   -1.2698     0.290     -4.382      0.000      -1.840     -0.70
age_75_more_2012                 -1.4948     0.365     -4.093      0.000      -2.213     -0.77
graduates_without_secondary_2012  0.5246     1.022      0.513      0.608      -1.485      2.53
graduates_lower_secondary_2012    0.5494     1.024      0.537      0.592      -1.464      2.56
graduates_secondary_2012          0.6049     1.021      0.593      0.554      -1.403      2.61
graduates_uni_2012                0.5730     1.021      0.561      0.575      -1.435      2.58
mining_manuf_2012                 1.3775     0.516      2.670      0.008       0.363      2.39
insolvencies_2012                 0.1412     0.222      0.637      0.525      -0.295      0.57
empl_manuf_2012                  -0.3827     0.260     -1.473      0.142      -0.893      0.12
```

```
empl_com_hotel_2012          -0.5137      0.268     -1.914      0.056      -1.042       0.01
empl_service_2012            -0.4151      0.260     -1.596      0.111      -0.927       0.09
empl_oth_service_2012        -0.3649      0.261     -1.400      0.162      -0.877       0.14
unempl_total_2013            -1.0337      0.401     -2.581      0.010      -1.821      -0.24
unempl_male_2013              1.1860      0.392      3.029      0.003       0.416       1.95
==============================================================================
Omnibus:                        6.246   Durbin-Watson:                   1.819
Prob(Omnibus):                  0.044   Jarque-Bera (JB):                6.522
Skew:                           0.221   Prob(JB):                       0.0384
Kurtosis:                       3.441   Cond. No.                     1.00e+16
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 8.61e-26. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

```
C:\Users\SOURAV CH\Anaconda3\lib\site-packages\numpy\core\fromnumeric.py:2389: FutureWarning: Meth
od .ptp is deprecated and will be removed in a future version. Use numpy.ptp instead.
  return ptp(axis=axis, out=out, **kwargs)
```

In [17]:

```
X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size = 0.3, random_state = 2)
reg = LinearRegression()
m3 = reg.fit(X_train, Y_train)
#Predict the model on train
pred_train = m3.predict(X_train)
print("MSE on train: ", mean_squared_error(Y_train,pred_train))
print("RSquared: ",m3.score(X_train,Y_train))
#Predict on test
pred_test = m3.predict(X_test)
print("MSE on test: ", mean_squared_error(Y_test,pred_test))
print("RSquared: ",m3.score(X_test,Y_test))
```

```
MSE on train:  2.121532024377327
RSquared:  0.9266704176621758
MSE on test:  2.6026600970093887
RSquared:  0.8991029374086554
```

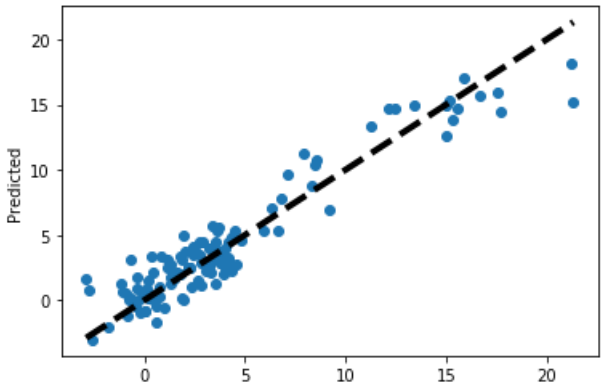
In [ ]:

```

```

- This might be the best model as the scores suggest it did well on MSe stats as well as the R^2 stats and also the prediction below were fair enough , i would chosse this feature over a over learning model.

In [249]:

```
fig, ax = plt.subplots()
ax.scatter(Y_test, pred_test)
ax.plot([Y.min(), Y.max()], [Y.min(), Y.max()], 'k--', lw=4)
ax.set_xlabel('Measured')
ax.set_ylabel('Predicted')
plt.show()
```



Measured

In [248]:

```
output = pd.DataFrame(pred_test,Y_test,)
output
```

Out[248]:

vot19_14	0
1.762276	3.421405
12.436741	14.674870
1.762954	1.946727
3.694157	5.564009
3.027529	2.302756
1.319050	1.295900
4.514345	5.390341
8.472181	10.389270
3.185247	2.593146
7.942587	11.313399
-0.386746	1.705652
4.207446	4.434960
17.741770	14.468807
1.895436	3.122920
-2.899875	1.650627
12.114440	14.665437
1.137629	2.529810
0.360833	3.360205
-0.728325	3.059134
3.535368	1.300015
-0.395556	0.835237
4.346890	4.786379
0.609086	-1.635894
0.551817	-0.501085
3.475091	3.145107
6.652091	5.305238
2.396875	3.800095
3.124034	2.932013
3.677925	3.466290
2.140122	2.543653
...	...
0.160638	0.234240
1.921189	4.915851
-0.481812	-0.484339
1.366896	2.145843
3.168329	2.860579
0.149390	0.110983
0.948635	-0.537281
1.952564	0.014442
2.524834	4.081914
0.751892	1.007971

4.193089	2.970592
8.565292	10.737535
2.812249	1.125089
1.288777	2.705431
8.346655	8.747109
21.251311	18.179521
0.576312	0.459109
3.421032	3.953887
6.761899	7.792657
-1.808341	-2.053493
2.391247	3.438978
3.906186	1.971982
2.883327	4.493312
-0.329070	-0.015836
3.404176	5.642970
1.590523	2.158971
4.188082	3.279752
0.179041	1.530986
0.766647	0.249986
0.028835	-0.837450

121 rows × 1 columns

Model 4 : using constant Feature elimination

In this mmodel we basically set a thresold of Variance of diff columns with respect to the target so based on thaat it eliminates all the features below this range.

- But this model might fail in test data.

In [6]:

```
df6 = df1.copy()

from sklearn.model_selection import train_test_split
from sklearn.feature_selection import VarianceThreshold
```

In [7]:

```
df6.drop(columns = {'subregion',"Nr"}, inplace = True)
```

In [10]:

```
X = df6.drop("vot19_14", axis = 1)
Y = df6["vot19_14"]
X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size = 0.3, random_state = 10)
constant_filter = VarianceThreshold(threshold=0.5)
constant_filter.fit(X_train)
```

Out[10]:

```
VarianceThreshold(threshold=0.5)
```

In [11]:

```
constant_columns = [column for column in X_train.columns
                    if column not in X_train.columns[constant_filter.get_support()]]

print(len(constant_columns))
```

6

In [253]:

```
for column in constant_columns:
    print(column)
```

```
turnout14
turnout19
turnout19_14
ove18_13
mining_manuf_2012
insolvencies_2012
```

In [254]:

```
X_train = constant_filter.transform(X_train)
X_test = constant_filter.transform(X_test)

X_train.shape, X_test.shape
```

Out[254]:

```
((280, 141), (121, 141))
```

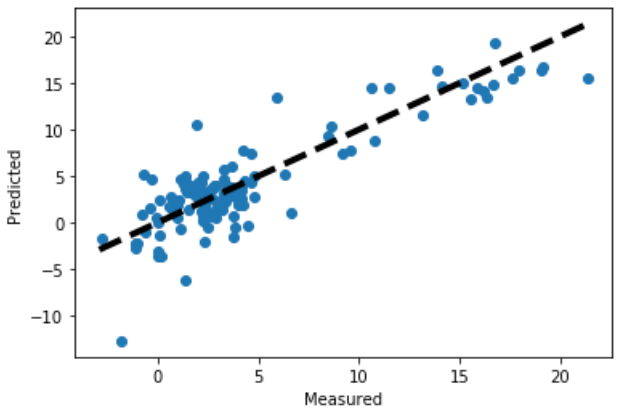
In [255]:

```
X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size = 0.3, random_state = 10)
reg = LinearRegression()
m4 = reg.fit(X_train, Y_train)
#Predict the model on train
pred_train = m4.predict(X_train)
print("MSE on train: ", mean_squared_error(Y_train,pred_train))
print("RSquared: ",m4.score(X_train,Y_train))
#Predict on test
pred_test = m4.predict(X_test)
print("MSE on test: ", mean_squared_error(Y_test,pred_test))
print("RSquared: ",m4.score(X_test,Y_test))
```

```
MSE on train:  1.0913312148592678
RSquared:  0.9616670989671317
MSE on test:  7.255112880255219
RSquared:  0.73243246686298
```

In [256]:

```
fig, ax = plt.subplots()
ax.scatter(Y_test, pred_test)
ax.plot([Y.min(), Y.max()], [Y.min(), Y.max()], 'k--', lw=4)
ax.set_xlabel('Measured')
ax.set_ylabel('Predicted')
plt.show()
```



Model 5 : Using highly correlated features

- Below is the model where we considered an absolute correlation with respect to the target , we only applied the model on correlated features that explain the model well
- As we know that cov(X,Y) =~ means the change in indep(X) var gives a change in dep(Y) variable it may be positively or vice cera

In [101]:

```
X = df2.drop("vot19_14", axis = 1)
y = df2["vot19_14"]
X_train, X_test, y_train, y_test = train_test_split(X,y, test_size = 0.3, random_state = 10)
reg = LinearRegression()
m5 = reg.fit(X_train, y_train)
#Predict the model on train
pred_train = m5.predict(X_train)
print("MSE on train: ", mean_squared_error(y_train,pred_train))
print("RSquared: ",m5.score(X_train,y_train))
#Predict on test
pred_test = m5.predict(X_test)
print("MSE on test: ", mean_squared_error(y_test,pred_test))
print("RSquared: ",m5.score(X_test,y_test))
```

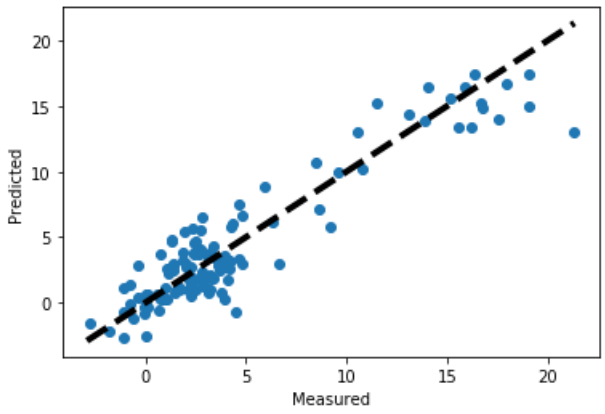
```
MSE on train:  3.4542165356646835
RSquared:  0.8786709856688132
MSE on test:  3.7889891592366176
RSquared:  0.8602626176666476
```

In [85]:

```
from sklearn.compose import TransformedTargetRegressor
```

In [102]:

```
fig, ax = plt.subplots()
ax.scatter(y_test, pred_test)
ax.plot([y.min(), y.max()], [y.min(), y.max()], 'k--', lw=4)
ax.set_xlabel('Measured')
ax.set_ylabel('Predicted')
plt.show()
```



log transformed model:

In [86]:

```
df8 = df1.copy()
```

In [87]:

```
df8["vot19_14_log"] = np.log(df8["vot19_14"])
```

```
C:\Users\SOURAV CH\Anaconda3\lib\site-packages\pandas\core\series.py:853: RuntimeWarning: invalid
value encountered in log
  result = getattr(ufunc, method)(*inputs, **kwargs)
```

In [88]:

```
df8.fillna(df8.median(),inplace = True)
df8.drop(columns = {"vot19_14"},inplace = True)
```

In [89]:

```
df8.isnull().sum().sort_values(ascending = False)
```

Out[89]:

```
vot19_14_log
unempl_total_2019
hartz_no_empl_2018
hartz_total_2018
empl_oth_service_2018
empl_service_2018
empl_com_hotel_2018
empl_manuf_2018
empl_agr_2018
empl_total_2018
insolvencies_2017
business_reg_2017
child_day_care_2018
graduates_higher_2017
graduates_secondary_2017
graduates_lower_secondary_2017
0
graduates_without_secondary_2017
0
hartz_foreign_2018
unempl_male_2019
f_crime_2012
unempl_female_2019
graduates_secondary_2017_2012
graduates_lower_secondary_2017_2012
0
graduates_without_secondary_2017_2012
0
age_75_more_2017_2012
ag4_60_74_2017_2012
age_35_59_2017_2012
age_25_34_2017_2012
age_18_24_2017_2012
age_to_18_2017_2012
net_migration_2017_2012
birth_balance_2017_2012
population_density_2017_2012
foreigners_2017_2012
unempl_55_64_2019
unempl_15_19_2019
Absolventen/Abgänger allgemeinbildender Schulen 2017 - insgesamt ohne Externe (je 1000 Einwohner)
0
graduates_voc_2017
vehicles_2018
space_per_inh_2017
foreigners_2017
germans_2017
population_2017
area_2017
ove18_13
debt_2018
debt_2017
debt_2016
debt_2015
debt_2014
debt_2013
turnout19_14
turnout19
turnout14
```

turnout14
subregion
population_density_2017
birth_balance_2017
net_migration_2017
protection_total_2017
space_per_app_2017
dwellings_2017
dwellings_new_2017
protection_rejected_2017
protection_accepted_2017
protection_open_2017
gdp_2016
age_to_18_2017
disposable_inc_2016
age_75_more_2017
ag4_60_74_2017
age_35_59_2017
age_25_34_2017
age_18_24_2017
dwellings_new_2017_2012
dwellings_2017_2012
vehicles_2018_2013
business_reg_2012
foreign_suspects_2018
total_suspects_2018
hartz_no_empl_2013
hartz_total_2013
unempl_female_2013
unempl_male_2013
unempl_total_2013
empl_oth_service_2012
empl_service_2012
empl_com_hotel_2012
empl_manuf_2012
empl_agr_2012
empl_soc_sec_total_2012
insolvencies_per_1000_2012
insolvencies_2012
f_crime_2018
total_suspects_2017
foreign_suspects_2017
foreign_suspects_2014
foreign_suspects_2012
total_suspects_2012
f_crime_2013
foreign_suspects_2013
total_suspects_2013
f_crime_2014
total_suspects_2014
f_crime_2017
f_crime_2015
foreign_suspects_2015
total_suspects_2015
f_crime_2016
foreign_suspects_2016
total_suspects_2016
business_delist_2012
trade_tax_per_inh_2012
business_reg_2017_2012
mining_manuf_2012
male_2012
population_2012
area_2012
unempl_female_2019_2013
unempl_male_2019_2013
unempl_total_2019_2013
hartz_no_empl_2018_2013
hartz_total_2018_2013
empl_oth_service_2018_2012
empl_service_2018_2012
empl_com_hotel_2018_2012
empl_manuf_2018_2012
empl_agr_2018_2012
empl_total_2018_2012
insolvencies_2017_2012
foreigners_2012
population_density_2012

```
population_density_2012
birth_balance_2012
graduates_without_secondary_2012
0
dwellings_2012
dwellings_new_2012
vehicles_2013
graduates_uni_2012
graduates_secondary_2012
graduates_lower_secondary_2012
0
graduates_sec_2012
net_migration_2012
age_75_more_2012
ag4_60_74_2012
age_35_59_2012
age_25_34_2012
age_18_24_2012
age_to_18_2012
Nr
dtype: int64
```

In [90]:

```
X = df8.drop("vot19_14_log", axis = 1)
Y = df8["vot19_14_log"]

X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size = 0.3, random_state = 2)
reg = LinearRegression()
m6 = reg.fit(X_train, Y_train)
#Predict the model on train
pred_train = m6.predict(X_train)
print("MSE on train: ", mean_squared_error(Y_train,pred_train))
print("RSquared: ",m6.score(X_train,Y_train))
#Predict on test
pred_test = m6.predict(X_test)
print("MSE on test: ", mean_squared_error(Y_test,pred_test))
print("RSquared: ",m6.score(X_test,Y_test))
```

```
MSE on train:  0.22660811434655595
RSquared:  0.8103182575219241
MSE on test:  1.095344969974244
RSquared:  0.12071484898622765
```

From the above model we can see that if the data is log tranformed inour case is got a good R^2 on the training data , but it greatly failed on the test data , so not always the log transformations are good.

In [ ]:

```

```