

# 403 Project

April 15, 2021

```
[1]: !pip install cjklib3
```

```
Collecting cjklib3
  Using cached cjklib3-0.0.3-py3-none-any.whl (28.1 MB)
Requirement already satisfied: SQLAlchemy<2.0,>=0.9.4 in
/opt/conda/lib/python3.8/site-packages (from cjklib3) (1.3.20)
Installing collected packages: cjklib3
Successfully installed cjklib3-0.0.3
```

```
[2]: #import sys
#!{sys.executable} -m pip install cjklib
import numpy as np
import pandas as pd
import cjklib
from cjklib import characterlookup
```

```
[125]: df = pd.read_csv('Chinese Lexicon Project Sze et al.csv')
```

```
[4]: # df
```

```
[133]: # Import the xinhua Chinese dictionary
xinhua = pd.read_csv("xinhua.csv")
```

```
[134]: char = list(df.Character)
rt = list(df.RT)
xinhua_char = list(xinhua.character)
radical = list(xinhua.radical)
```

```
[135]: ##### Study 1 #####
# Correlation between stroke counts and rt
# check stroke counts
charlookup = characterlookup.CharacterLookup('T')
stroke_count = []
for i in char:
    stroke_count.append(charlookup.getStrokeCount(i))

# Uncomment to check the stroke count
```

```
#stroke_count
```

```
[136]: ##### Study 2 #####  
# Correlation between radical_removed stroke counts and rt  
# get the radical for each character  
new_char = []  
radical_list = []  
for i in range(len(char)):  
    for j in range(len(xinhua_char)):  
        if char[i] == xinhua_char[j]:  
            radical_list.append(radical[j])  
            new_char.append(char[i])  
  
# print(radical_list)  
# print(new_char)
```

```
[137]: # Check missing characters  
for i in char:  
    if i not in new_char:  
        print(i)
```

```
[138]: # get the stroke count for each charcter in the new list  
new_stroke_count = []  
for i in new_char:  
    new_stroke_count.append(charlookup.getStrokeCount(i))  
  
# get the radical count for each charcter  
radical_count = []  
for j in radical_list:  
    radical_count.append(charlookup.getStrokeCount(j))  
  
#print(radical_count)
```

```
[139]: # check radical removed stroke counts  
radical_removed = []  
for i in range(len(new_stroke_count)):  
    x = new_stroke_count[i] - radical_count[i]  
    radical_removed.append(x)  
  
#print(radical_removed)
```

```
[140]: # get the reaction time list for radical removed characters
new_rt = []
for i in range(len(char)):
    for j in range(len(new_char)):
        if char[i] == new_char[j]:
            new_rt.append(rt[i])

# new_rt
```

```
[141]: # Check nulls for radical removed characters
for i in range(len(radical_removed)):
    if radical_removed[i] == 0:
        print(new_char[i])
```

```
[142]: ##### Study 3 #####
# Compare rt for radical_removed characters within each radical.

# Get all the radicals used in the study
radicals = []
for k in radical_list:
    if k not in radicals:
        radicals.append(k)

# Get the dictionary containing all characters for each radical.
radical_dict = {}
for k in radicals:
    l = []
    for i in range(len(new_char)):
        if radical_list[i] == k:
            l.append(new_char[i])
    radical_dict[k] = l
```

```
[143]: print(radical_dict)
```







```
[144]: len(radicals)
```

```
[145]: ##### Data Analysis #####
```

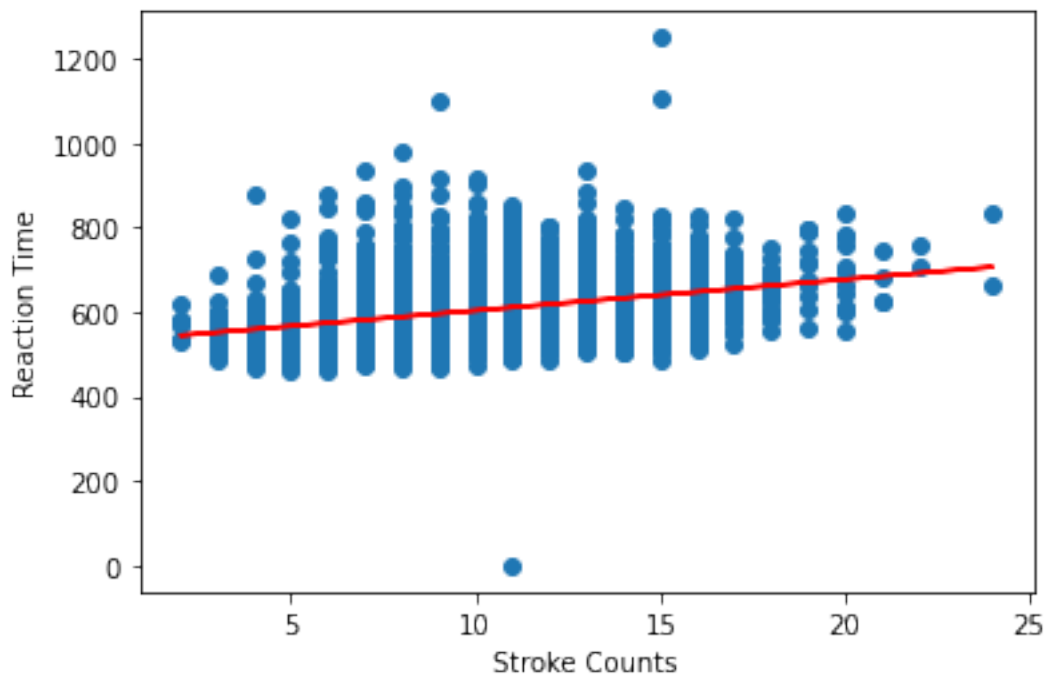
```
[146]: # Study 1 linear regression results
slope1, intercept1 = np.polyfit(stroke_count, rt, 1)
result_1 = stats.linregress(stroke_count, rt)
print(result_1)
```

7

```
rvalue=0.3098711047884425, pvalue=8.96616586439971e-57,
stderr=0.4524448772607573)
```

```
[147]: # Study 1 result graph
plt.scatter(stroke_count, rt)
plt.plot(stroke_count, np.multiply(slope1, stroke_count) + intercept1, 'r-')
plt.xlabel("Stroke Counts")
plt.ylabel("Reaction Time")
```

```
[147]: Text(0, 0.5, 'Reaction Time')
```



```
[148]: # Study 2 linear regression results
slope2, intercept2 = np.polyt(radical_removed, new_rt, 1)
result_2 = stats.linregress(radical_removed, new_rt)
print(result_2)
```

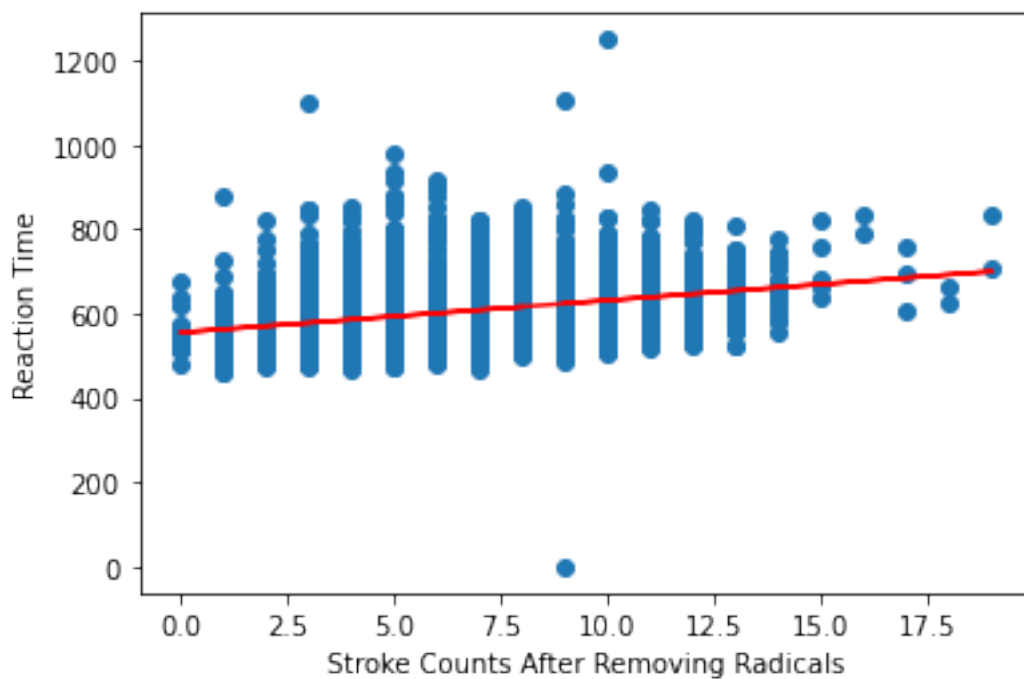
```
LinregressResult(slope=7.619099272309819, intercept=555.0624164800901,
rvalue=0.2790367734354089, pvalue=7.186914042240392e-46,
stderr=0.5250395735895467)
```

```
[149]: # Study 2 result graph
plt.scatter(radical_removed, new_rt)
plt.plot(radical_removed, np.multiply(slope2, radical_removed) + intercept2, 'r-')
plt.xlabel("Stroke Counts After Removing Radicals")
```



```
plt.ylabel("Reaction Time")
```

```
[149]: Text(0, 0.5, 'Reaction Time')
```



```
[155]: # Study 3 analysis
r_value_list = []
p_value_list = []
for j in radicals:
    r = []
    for i in range(len(new_rt)):
        if radical_list[i] == j:
            r.append(new_rt[i])

    s = []
    for i in range(len(new_char)):
        if radical_list[i] == j:
            s.append(radical_removed[i])

    result_3 = stats.linregress(s, r)
    slope, intercept, r_value, p_value, std_err = result_3
    r_value_list.append(r_value)
    if r_value != 0:
        p_value_list.append(p_value)
    print("Radical:", j, "\n", result_3)
```

Radical:  
 LinregressResult(slope=6.866538857920729, intercept=570.9566750089184, rvalue=0.29127260754649353, pvalue=0.0005282204588124549, stderr=1.9338258558576813)

Radical:  
 LinregressResult(slope=2.4384992833995467, intercept=593.6891524446078, rvalue=0.1199680321952138, pvalue=0.4379400783954325, stderr=3.113755368655418)

Radical:  
 LinregressResult(slope=7.509133603189079, intercept=568.4829332075204, rvalue=0.30901253240780585, pvalue=0.0003291363990149997, stderr=2.0348194052223794)

Radical:  
 LinregressResult(slope=-2.5107147639771834, intercept=701.2227055027982, rvalue=-0.08074082881587238, pvalue=0.688903002778481, stderr=6.1988900593216725)

Radical:  
 LinregressResult(slope=11.605844777906993, intercept=511.30046188372086, rvalue=0.5442692819843777, pvalue=0.34293767202762837, stderr=10.328041498867051)

Radical:  
 LinregressResult(slope=7.538142083606431, intercept=564.6461205347191, rvalue=0.26726116153473306, pvalue=0.022264843975859663, stderr=3.225573641414224)

Radical:  
 LinregressResult(slope=-0.30038193962883347, intercept=598.1071036640399, rvalue=-0.008661351589170157, pvalue=0.9588398604078958, stderr=5.7799037703502645)

Radical:  
 LinregressResult(slope=23.012004799999996, intercept=358.98679470000013, rvalue=1.0, pvalue=0.0, stderr=0.0)

Radical:  
 LinregressResult(slope=-8.07652580502794, intercept=681.6233525593576, rvalue=-0.27362121962091734, pvalue=0.19574732913121162, stderr=6.052923535808962)

Radical:  
 LinregressResult(slope=15.71857219021325, intercept=479.79159023077693, rvalue=0.4830184929202601, pvalue=0.0007773135086778994, stderr=4.34536132904413)

Radical:  
 LinregressResult(slope=-7.993798117105261, intercept=698.5549244328947, rvalue=-0.561039340769764, pvalue=0.32514286966308226, stderr=6.809558525253224)

Radical:  
 LinregressResult(slope=4.8764901191606835, intercept=580.6948974548317, rvalue=0.09868334660541696, pvalue=0.48641704552982845, stderr=6.954300387738346)

Radical:  
 LinregressResult(slope=7.087000132694296, intercept=579.7271724245595, rvalue=0.32336442434018176, pvalue=0.1905583633935934,

```

stderr=5.1847444215591425)
Radical:
  LinregressResult(slope=11.230597380654633, intercept=538.6950755801745,
rvalue=0.46188909739732087, pvalue=0.0010822166347165007,
stderr=3.2147854882510494)
Radical:
  LinregressResult(slope=13.484997975000008, intercept=566.94848805,
rvalue=0.40564788274434066, pvalue=0.3665821659290805,
stderr=13.588666757156963)
Radical:
  LinregressResult(slope=3.609302183181125, intercept=565.8029767715374,
rvalue=0.12203908766098143, pvalue=0.39359145010305274, stderr=4.19341491616144)
Radical:
  LinregressResult(slope=2.654692163375, intercept=555.4849999484375,
rvalue=0.09656894236512689, pvalue=0.65351412727647, stderr=5.833521882216274)
Radical:
  LinregressResult(slope=2.060743870822171, intercept=606.4874635761695,
rvalue=0.06955245591961319, pvalue=0.46014139817627053,
stderr=2.7804770642202272)
Radical:
  LinregressResult(slope=3.9139437201541947, intercept=598.7683911152975,
rvalue=0.15753872452594964, pvalue=0.2695619745762643, stderr=3.504870286533772)
Radical:
  LinregressResult(slope=-6.501008400000001, intercept=544.1774790000001,
rvalue=-1.0, pvalue=0.0, stderr=0.0)
Radical:
  LinregressResult(slope=9.426100867901237, intercept=579.7870598022635,
rvalue=0.23178375631105683, pvalue=0.1737571021034979, stderr=6.78451187246574)
Radical:
  LinregressResult(slope=8.906140500868652, intercept=573.3112081388117,
rvalue=0.38941599132193555, pvalue=0.05998202664290961,
stderr=4.4911057300640005)
Radical:
  LinregressResult(slope=15.262605000000008, intercept=477.80672289999995,
rvalue=1.0, pvalue=0.0, stderr=0.0)
Radical:
  LinregressResult(slope=22.67927471584614, intercept=429.9775279709231,
rvalue=0.8739107692563404, pvalue=0.002067446924647713,
stderr=4.767893011785514)
Radical:
  LinregressResult(slope=11.894492442876343, intercept=504.64285119207,
rvalue=0.46683186834853146, pvalue=0.14773332078760484,
stderr=7.510803921928355)
Radical:
  LinregressResult(slope=34.82459108090909, intercept=476.4256621654545,
rvalue=0.5286631804687342, pvalue=0.22248149501344958,
stderr=25.005960746117154)
Radical:

```

```

LinregressResult(slope=8.434226865706316, intercept=547.2619738819703,
rvalue=0.36106143149384146, pvalue=0.33974596978506605, stderr=8.23348265411168)
Radical:
LinregressResult(slope=3.6636309669182476, intercept=556.8196443180971,
rvalue=0.1513598933732722, pvalue=0.3576757063402196, stderr=3.9333932753104075)
Radical:
LinregressResult(slope=6.542072332736572, intercept=563.9029159790282,
rvalue=0.21983993462991014, pvalue=0.35169155383984374,
stderr=6.842515412636792)
Radical:
LinregressResult(slope=-63.27393209999993, intercept=857.7125175999997,
rvalue=-0.3187513968568805, pvalue=0.6812486031431195,
stderr=133.04299359849412)
Radical:
LinregressResult(slope=-13.136363650000078, intercept=722.5151516000005,
rvalue=-1.0, pvalue=0.0, stderr=0.0)
Radical:
LinregressResult(slope=-11.856906946879645, intercept=687.126343172734,
rvalue=-0.20477557339873104, pvalue=0.4641064013764873,
stderr=15.718804675571311)
Radical:
LinregressResult(slope=12.493971493867177, intercept=528.9909444037529,
rvalue=0.45443480414698056, pvalue=3.729203939580291e-06,
stderr=2.5395566700365997)
Radical:
LinregressResult(slope=3.4732204836221823, intercept=549.8904848375216,
rvalue=0.12667656761158186, pvalue=0.6528019410857124, stderr=7.543129742089087)
Radical:
LinregressResult(slope=7.392178579701481, intercept=546.4953522414268,
rvalue=0.3161646155813427, pvalue=0.001800707719717451,
stderr=2.300109981639268)
Radical:
LinregressResult(slope=7.444974188801254, intercept=540.2724068004733,
rvalue=0.24378110402322983, pvalue=0.30031617353904666,
stderr=6.9810804711936765)
Radical:
LinregressResult(slope=7.949318781186289, intercept=532.6669361944394,
rvalue=0.21527918281345315, pvalue=0.4232935591847051, stderr=9.637390918189165)
Radical:
LinregressResult(slope=3.2119303217293917, intercept=611.4428850861558,
rvalue=0.1488998789566684, pvalue=0.4774815258107308, stderr=4.447738757311624)
Radical:
LinregressResult(slope=7.930795432837578, intercept=598.5334126082654,
rvalue=0.3137392253418648, pvalue=0.07540376970644624, stderr=4.310883165605697)
Radical:
LinregressResult(slope=-9.03232821176471, intercept=553.4187839588236,
rvalue=-0.14958457780354104, pvalue=0.7772966515892223,
stderr=29.851689994438924)

```

Radical:  
 LinregressResult(slope=3.2030395894310657, intercept=606.1500158109678, rvalue=0.08605617311524402, pvalue=0.5169391179441248, stderr=4.911665084683761)

Radical:  
 LinregressResult(slope=-3.415380966651628, intercept=597.6256720134927, rvalue=-0.08183450007987438, pvalue=0.6849003879205566, stderr=8.31904769928076)

Radical:  
 LinregressResult(slope=2.6693091750566373, intercept=604.8793382645221, rvalue=0.12555404571855586, pvalue=0.45903281856536593, stderr=3.5651992674520105)

Radical:  
 LinregressResult(slope=-0.7758510047468311, intercept=632.7066962531645, rvalue=-0.016584832578827192, pvalue=0.9613994243327613, stderr=15.591440597384304)

Radical:  
 LinregressResult(slope=-25.235294124999996, intercept=710.941176475, rvalue=-1.0, pvalue=0.0, stderr=0.0)

Radical:  
 LinregressResult(slope=-42.815126099999993, intercept=650.1596639999998, rvalue=-1.0, pvalue=0.0, stderr=0.0)

Radical:  
 LinregressResult(slope=26.362834109833187, intercept=524.0808328019316, rvalue=0.4053522532023764, pvalue=0.1911263867242264, stderr=18.80104510595725)

Radical:  
 LinregressResult(slope=-11.528074857142851, intercept=592.1192067357142, rvalue=-0.7405534385458086, pvalue=0.25944656145419126, stderr=7.396955868460947)

Radical:  
 LinregressResult(slope=0.843593340543729, intercept=574.5055733555556, rvalue=0.04165238829131558, pvalue=0.8782722320947062, stderr=5.408192679390734)

Radical:  
 LinregressResult(slope=14.02591202068015, intercept=528.7441592727022, rvalue=0.5094797978001628, pvalue=0.06275566504205049, stderr=6.838417650076792)

Radical:  
 LinregressResult(slope=9.017495933766227, intercept=536.9761840311688, rvalue=0.3242517270779255, pvalue=0.019020768770391504, stderr=3.72045795757331)

Radical:  
 LinregressResult(slope=4.245303862085862, intercept=551.4162391751983, rvalue=0.1772802275787273, pvalue=0.29386313202613223, stderr=3.9836424399698496)

Radical:  
 LinregressResult(slope=-3.829675425783973, intercept=600.4400825200349, rvalue=-0.15631024267059418, pvalue=0.5631997667489337, stderr=6.467539465876749)

Radical:  
 LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)

Radical:  
 LinregressResult(slope=6.417971593819336, intercept=528.9509764114106,

```

rvalue=0.3551993310206668, pvalue=0.12433353524307578, stderr=3.98110587785624)
Radical:
  LinregressResult(slope=28.886573791666667, intercept=451.46284774935896,
rvalue=0.5244711720369237, pvalue=0.06576280777547124,
stderr=14.139241178430428)
Radical:
  LinregressResult(slope=25.573485807575754, intercept=403.35197894242424,
rvalue=0.9212297670893859, pvalue=0.02622273465722521, stderr=6.234936820863288)
Radical:
  LinregressResult(slope=-5.113827336363639, intercept=555.3116882727272,
rvalue=-0.18738536476712472, pvalue=0.8126146352328752,
stderr=18.955426466384115)
Radical:
  LinregressResult(slope=1.3409909133233533, intercept=539.4916916297905,
rvalue=0.14150883943572892, pvalue=0.6446980792919084, stderr=2.828482373310844)
Radical:
  LinregressResult(slope=22.353023789583336, intercept=504.75648914583337,
rvalue=0.5058382921334118, pvalue=0.1124080994568579, stderr=12.706532045865162)
Radical:
  LinregressResult(slope=-8.606598626829268, intercept=638.3161765,
rvalue=-0.7590242476047159, pvalue=0.13675300439897262,
stderr=4.262239583583602)
Radical:
  LinregressResult(slope=1.9180668737581021, intercept=599.4077614176025,
rvalue=0.05837913527409738, pvalue=0.829957364461978, stderr=8.765985351858834)
Radical:
  LinregressResult(slope=3.3797579364780095, intercept=505.93414726540885,
rvalue=0.19716155656937004, pvalue=0.6397906059363276, stderr=6.860854265828319)
Radical:
  LinregressResult(slope=-6.041646527595636, intercept=563.9355866478143,
rvalue=-0.225048013845268, pvalue=0.3266872450632772, stderr=6.000912316681663)
Radical:
  LinregressResult(slope=6.359452556081939, intercept=575.7143375709164,
rvalue=0.23768644833568583, pvalue=0.22325288559483486,
stderr=5.096837349037842)
Radical:
  LinregressResult(slope=17.88168684210529, intercept=461.53265246315783,
rvalue=0.6591769245681439, pvalue=0.3408230754318561, stderr=14.424552085935634)
Radical:
  LinregressResult(slope=-3.1453453834578458, intercept=631.0589062329775,
rvalue=-0.11635861984998928, pvalue=0.6796311700738293,
stderr=7.446256487826656)
Radical:
  LinregressResult(slope=-1.4778076738461388, intercept=553.2357297338461,
rvalue=-0.0303269613086773, pvalue=0.9545235042628939,
stderr=24.353378409040126)
Radical:
  LinregressResult(slope=-1.6688965529467599, intercept=585.1280690225285,

```

```

rvalue=-0.061600432364370426, pvalue=0.814316372758636,
stderr=6.981913605918642)
Radical:
  LinregressResult(slope=19.606617650000004, intercept=538.3235294000001,
rvalue=1.0, pvalue=0.0, stderr=0.0)
Radical:
  LinregressResult(slope=-19.575139164285762, intercept=633.5142858000002,
rvalue=-0.9703036568435982, pvalue=0.15553468481924884, stderr=4.87994872003339)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=12.162169460465119, intercept=530.9721256627907,
rvalue=0.5165589057117246, pvalue=0.4834410942882754, stderr=14.255341414780464)
Radical:
  LinregressResult(slope=0.7884773019354762, intercept=541.3562247690323,
rvalue=0.021881816487849072, pvalue=0.9461845811242282,
stderr=11.392048525245428)
Radical:
  LinregressResult(slope=-9.657585548618783, intercept=623.3939514353591,
rvalue=-0.3440255964294202, pvalue=0.4499028640907149,
stderr=11.787997140270178)
Radical:
  LinregressResult(slope=4.883577268856439, intercept=561.6649569416059,
rvalue=0.11617338378952172, pvalue=0.6461931430447361,
stderr=10.438084351408593)
Radical:
  LinregressResult(slope=12.847836262908022, intercept=523.9280818030662,
rvalue=0.37152750708890236, pvalue=0.23439898426714834,
stderr=10.152766961022902)
Radical:
  LinregressResult(slope=18.40411565539651, intercept=466.2255214941199,
rvalue=0.6922336443570599, pvalue=0.001455194890847459,
stderr=4.796704810443728)
Radical:
  LinregressResult(slope=-4.9686323622806965, intercept=559.5743320192981,
rvalue=-0.37204703105097914, pvalue=0.5374622422008264,
stderr=7.156924757911497)
Radical:
  LinregressResult(slope=-11.732254153979234, intercept=623.1671853044983,
rvalue=-0.26795385993572834, pvalue=0.45415399405066326,
stderr=14.914112131658642)
Radical:
  LinregressResult(slope=3.38770691327201, intercept=574.3723331627135,
rvalue=0.15750532996397443, pvalue=0.33172779600433233,
stderr=3.4455920601373684)
Radical:
  LinregressResult(slope=20.329411779999987, intercept=512.4823528600002,
rvalue=1.0, pvalue=0.0, stderr=0.0)

```

Radical:  
 LinregressResult(slope=1.7932387418495348, intercept=596.6401575521943, rvalue=0.0443944177648801, pvalue=0.8656527770630783, stderr=10.4192324911014)

Radical:  
 LinregressResult(slope=-1.7921903329431381, intercept=592.1722861981605, rvalue=-0.3539502355295238, pvalue=0.4360367746272024, stderr=2.1178307993306955)

Radical:  
 LinregressResult(slope=-1.1517507035714258, intercept=555.8361344571429, rvalue=-0.18976892219397754, pvalue=0.7598368611514329, stderr=3.4403967346707627)

Radical:  
 LinregressResult(slope=5.517532530631863, intercept=560.4365354607144, rvalue=0.3459449130003771, pvalue=0.3274942847685285, stderr=5.2907076788802625)

Radical:  
 LinregressResult(slope=-3.9831288599303107, intercept=595.435374625087, rvalue=-0.1543100105523192, pvalue=0.6505377222072645, stderr=8.501113741902953)

Radical:  
 LinregressResult(slope=-19.044851723076935, intercept=675.5430584461539, rvalue=-0.6330299691630027, pvalue=0.09203903598881052, stderr=9.508019771446367)

Radical:  
 LinregressResult(slope=-14.664355810989017, intercept=640.8789378604396, rvalue=-0.8222192069930873, pvalue=0.04459954475026477, stderr=5.075608381392427)

Radical:  
 LinregressResult(slope=1.3406150648648656, intercept=581.0878557912162, rvalue=0.09816954725064907, pvalue=0.8171112865928042, stderr=5.548158103226752)

Radical:  
 LinregressResult(slope=9.925910428399996, intercept=503.2612578424001, rvalue=0.5516043781801047, pvalue=0.07856009986308557, stderr=5.003139040812031)

Radical:  
 LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)

Radical:  
 LinregressResult(slope=-7.178003118989897, intercept=600.7391058438384, rvalue=-0.32821627640892326, pvalue=0.4273628930454938, stderr=8.433678494421295)

Radical:  
 LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=0.0, stderr=0.0)

Radical:  
 LinregressResult(slope=15.393055884770789, intercept=516.9737915109556, rvalue=0.6790826205549267, pvalue=7.092024658466835e-05, stderr=3.263233930226811)

Radical:  
 LinregressResult(slope=14.308080799999997, intercept=509.5, rvalue=1.0, pvalue=0.0, stderr=0.0)

Radical:  
 LinregressResult(slope=25.08964803103447, intercept=461.94247416206895,



```

rvalue=0.6399111772341008, pvalue=0.03396238024946753,
stderr=10.043118837702801)
Radical:
  LinregressResult(slope=-6.590201338888893, intercept=574.9546348111111,
rvalue=-0.42364532153043316, pvalue=0.4771978436992507, stderr=8.13544765364078)
Radical:
  LinregressResult(slope=17.357358850000026, intercept=636.9093638999998,
rvalue=0.43848754959292574, pvalue=0.7110286598892824, stderr=35.57616449859805)
Radical:
  LinregressResult(slope=-2.7960183969387713, intercept=626.0322033520408,
rvalue=-0.07369938927581933, pvalue=0.8505387569691654,
stderr=14.300276798630689)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=0.0, stderr=0.0)
Radical:
  LinregressResult(slope=16.15445317905759, intercept=526.149147591623,
rvalue=0.47222988700579327, pvalue=0.23739726799368008,
stderr=12.310434406880642)
Radical:
  LinregressResult(slope=-3.2229978348484836, intercept=581.6524002662338,
rvalue=-0.32639123114141066, pvalue=0.47493765329088405,
stderr=4.174228473764214)
Radical:
  LinregressResult(slope=10.122280530223886, intercept=513.0734871026119,
rvalue=0.4024938584027476, pvalue=0.24885013789452068, stderr=8.139463663325088)
Radical:
  LinregressResult(slope=3.3469532923645353, intercept=622.9956917408866,
rvalue=0.2018092223899486, pvalue=0.5517988378015973, stderr=5.41450185667388)
Radical:
  LinregressResult(slope=14.603066474509799, intercept=509.6071348450981,
rvalue=0.8435435451383849, pvalue=0.00029026456919655625,
stderr=2.8032617802189916)
Radical:
  LinregressResult(slope=0.9515339131578875, intercept=521.3468162263158,
rvalue=0.09713271719121366, pvalue=0.8547591363812421, stderr=4.874951269855757)
Radical:
  LinregressResult(slope=60.442190650000004, intercept=284.8762678999997,
rvalue=1.0, pvalue=0.0, stderr=0.0)
Radical:
  LinregressResult(slope=7.43381193884893, intercept=503.3236380679856,
rvalue=0.49088122390384653, pvalue=0.14968232981229582,
stderr=4.664673332350274)
Radical:
  LinregressResult(slope=16.389627962931034, intercept=467.4444341362069,
rvalue=0.5224464878552904, pvalue=0.0992028265829452, stderr=8.91636968696343)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:

```

```

LinregressResult(slope=3.505494287500007, intercept=559.2301831875,
rvalue=0.5924728621133739, pvalue=0.29245093583598003,
stderr=2.7519114872772916)
Radical:
LinregressResult(slope=13.637178831864412, intercept=450.7909241674576,
rvalue=0.4707634903029983, pvalue=0.08932458615046145, stderr=7.377809837711642)
Radical:
LinregressResult(slope=6.828071923076909, intercept=548.4197802461539,
rvalue=0.17450662268622463, pvalue=0.8883339124012183, stderr=38.52748901471519)
Radical:
LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
LinregressResult(slope=84.95187170000003, intercept=307.5080212999999,
rvalue=1.0, pvalue=0.0, stderr=0.0)
Radical:
LinregressResult(slope=13.438324532203362, intercept=550.4696579915255,
rvalue=0.4128894409178999, pvalue=0.3572441662344023, stderr=13.25685119964008)
Radical:
LinregressResult(slope=38.20145245000003, intercept=500.8090779499998,
rvalue=0.7244566849716928, pvalue=0.4841814535846583, stderr=36.34867093671928)
Radical:
LinregressResult(slope=-2.4140983879581257, intercept=575.5670544607331,
rvalue=-0.1715945494325588, pvalue=0.6845201160590799, stderr=5.658300358170183)
Radical:
LinregressResult(slope=3.4656443645985426, intercept=509.28800216532846,
rvalue=0.5030549524046644, pvalue=0.38763709148602143, stderr=3.437554174465529)
Radical:
LinregressResult(slope=8.445946805952378, intercept=550.5131143531746,
rvalue=0.6219890069044149, pvalue=0.07369398268765653,
stderr=4.0187665707021045)
Radical:
LinregressResult(slope=14.17532182128378, intercept=497.35195386925676,
rvalue=0.7639920637511161, pvalue=0.04555442489430284, stderr=5.353905423611498)
Radical:
LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
LinregressResult(slope=1.1692159777777913, intercept=574.3408368666666,
rvalue=1.0, pvalue=0.0, stderr=0.0)
Radical:
LinregressResult(slope=-5.248147092857138, intercept=660.8989677000001,
rvalue=-0.21081071995630757, pvalue=0.8647792585957604,
stderr=24.335598112995623)
Radical:
LinregressResult(slope=9.383278903246755, intercept=556.5131653681818,
rvalue=0.3384607502837391, pvalue=0.5116951751878689, stderr=13.043584170439983)
Radical:

```

```

LinregressResult(slope=0.7041751329383948, intercept=647.73028659218,
rvalue=0.07704423928035757, pvalue=0.9509034606614601, stderr=9.11271371244638)
Radical:
LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
LinregressResult(slope=3.8968254000000306, intercept=673.3333332999998,
rvalue=1.0, pvalue=0.0, stderr=0.0)
Radical:
LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=1.0, stderr=inf)
Radical:
LinregressResult(slope=23.58444936989794, intercept=457.9377880540818,
rvalue=0.8487461532744811, pvalue=0.032586417181695346,
stderr=7.346995646163173)
Radical:
LinregressResult(slope=-40.79411760000005, intercept=738.3823528000003,
rvalue=-1.0, pvalue=0.0, stderr=0.0)
Radical:
LinregressResult(slope=45.68548227882354, intercept=375.54976178705874,
rvalue=0.6735893016405863, pvalue=0.04668998357723577,
stderr=18.947008541267056)
Radical:
LinregressResult(slope=31.440340262765943, intercept=384.62851731595754,
rvalue=0.9606120512062863, pvalue=0.00932816568886056, stderr=5.251170210710139)
Radical:
LinregressResult(slope=1.1351775451219466, intercept=546.1650152475611,
rvalue=0.09916395610856682, pvalue=0.851741629732002, stderr=5.695529009900589)
Radical:
LinregressResult(slope=28.251793131250007, intercept=450.1147468294643,
rvalue=0.8899978307402745, pvalue=0.007258919560573506,
stderr=6.472966284407239)
Radical:
LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
LinregressResult(slope=-32.601508620000004, intercept=803.21831896,
rvalue=-1.0, pvalue=0.0, stderr=0.0)
Radical:
LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
LinregressResult(slope=-28.86174243333331, intercept=776.7594697333332,
rvalue=-1.0, pvalue=0.0, stderr=0.0)
Radical:
LinregressResult(slope=6.334603580736539, intercept=557.1497837328612,
rvalue=0.4832603556747522, pvalue=0.33153991627949536, stderr=5.737898513848271)
Radical:
LinregressResult(slope=16.96503268888889, intercept=462.1941176,
rvalue=0.8734861885889696, pvalue=0.12651381141103032, stderr=6.686177139213768)
Radical:
LinregressResult(slope=-13.021217879411763, intercept=673.3521591941176,

```

```

rvalue=-0.313880419841454, pvalue=0.6070178316485689, stderr=22.74073935664201)
Radical:
  LinregressResult(slope=-5.396066329999998, intercept=552.31205595,
rvalue=-0.2761235232730869, pvalue=0.5079824821138597, stderr=7.667907004258953)
Radical:
  LinregressResult(slope=-6.996498600000019, intercept=598.2019141000002,
rvalue=-0.6387837167113128, pvalue=0.5588757333975494, stderr=8.426968964315202)
Radical:
  LinregressResult(slope=115.63963130000002, intercept=137.73041489999997,
rvalue=0.9340990540641886, pvalue=0.2324102780865131, stderr=44.19762640072307)
Radical:
  LinregressResult(slope=18.412180545177645, intercept=488.4005917649747,
rvalue=0.8602621293161081, pvalue=0.02792570093632795, stderr=5.456178344421962)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=-3.004648097222207, intercept=635.0830479416666,
rvalue=-0.07614607069928528, pvalue=0.9031415850178264,
stderr=22.715523876820225)
Radical:
  LinregressResult(slope=61.38787879999995, intercept=477.0121212, rvalue=1.0,
pvalue=0.0, stderr=0.0)
Radical:
  LinregressResult(slope=-95.7058823000001, intercept=1095.5882350000004,
rvalue=-1.0, pvalue=0.0, stderr=0.0)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=17.85642767765957, intercept=482.7719113138297,
rvalue=0.9872069325748074, pvalue=0.0017336472700608127,
stderr=1.6650782445352361)
Radical:
  LinregressResult(slope=16.115754109999997, intercept=538.4437984900002,
rvalue=0.47479780015647255, pvalue=0.3413208343820174,
stderr=14.936245643219925)
Radical:
  LinregressResult(slope=2.9121219428571505, intercept=615.7566917285715,
rvalue=0.14995105360432026, pvalue=0.8500489463956797,
stderr=13.577089528313474)
Radical:
  LinregressResult(slope=6.68899711851851, intercept=570.2677272333334,
rvalue=0.39750145049603414, pvalue=0.5075489713306067, stderr=8.914886129420912)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=19.2643428192308, intercept=519.6361889115382,
rvalue=0.4757263916529284, pvalue=0.6843704904840782, stderr=35.61876630005387)
Radical:

```

```

LinregressResult(slope=-23.761471900000004, intercept=595.8943724000001,
rvalue=-1.0, pvalue=0.0, stderr=0.0)
Radical:
LinregressResult(slope=32.223809549999998, intercept=412.71428560000004,
rvalue=0.9785251097217447, pvalue=0.13217243100596188, stderr=6.7879895893484)
Radical:
LinregressResult(slope=-46.85149188913042, intercept=739.5205455717391,
rvalue=-0.47087755992760066, pvalue=0.42341807621695726,
stderr=50.67819687838662)
Radical:
LinregressResult(slope=30.185714299999972, intercept=434.91428570000005,
rvalue=1.0, pvalue=0.0, stderr=0.0)
Radical:
LinregressResult(slope=56.470588225, intercept=90.97058832499988, rvalue=1.0,
pvalue=0.0, stderr=0.0)
Radical:
LinregressResult(slope=29.383943708474597, intercept=449.5421004762711,
rvalue=0.7823677477468396, pvalue=0.21763225225316027, stderr=16.54033999398382)
Radical:
LinregressResult(slope=2.0551020285714356, intercept=528.5122449285715,
rvalue=0.15469777324066447, pvalue=0.9011192386827628,
stderr=13.124703353687963)
Radical:
LinregressResult(slope=2.937846352264799, intercept=528.4613664986064,
rvalue=0.17491773789508575, pvalue=0.6786576128016895, stderr=6.751059828844368)
Radical:
LinregressResult(slope=47.371682209999975, intercept=268.5427819700002,
rvalue=0.8590097201748432, pvalue=0.14099027982515672, stderr=19.96370518290595)
Radical:
LinregressResult(slope=24.279875983146052, intercept=393.0978315651687,
rvalue=0.9709965561328974, pvalue=0.0012496007892739708,
stderr=2.989279309403978)
Radical:
LinregressResult(slope=-2.892156866666672, intercept=564.7352941, rvalue=-1.0,
pvalue=0.0, stderr=0.0)
Radical:
LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
LinregressResult(slope=67.159523799999996, intercept=428.8976191,
rvalue=0.9999999999999998, pvalue=0.0, stderr=0.0)
Radical:
LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=0.0, stderr=0.0)
Radical:
LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)

```

```

Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=2.4191176499999756, intercept=517.2500000000001,
rvalue=1.0, pvalue=0.0, stderr=0.0)
Radical:
  LinregressResult(slope=-0.03116885000002867, intercept=569.0961040000002,
rvalue=-1.0, pvalue=0.0, stderr=0.0)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=-21.085064950000003, intercept=683.70974035,
rvalue=-1.0, pvalue=0.0, stderr=0.0)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=-14.264705900000024, intercept=666.7058825000001,
rvalue=-1.0, pvalue=0.0, stderr=0.0)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=57.746130040000004, intercept=419.5665634200001,
rvalue=1.0, pvalue=0.0, stderr=0.0)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)
Radical:
  LinregressResult(slope=nan, intercept=nan, rvalue=0.0, pvalue=nan, stderr=nan)

```

```

[151]: ave_r = sum(r_value_list)/len(r_value_list)
print("Average Pearson's r correlation coefficient is: ", ave_r)

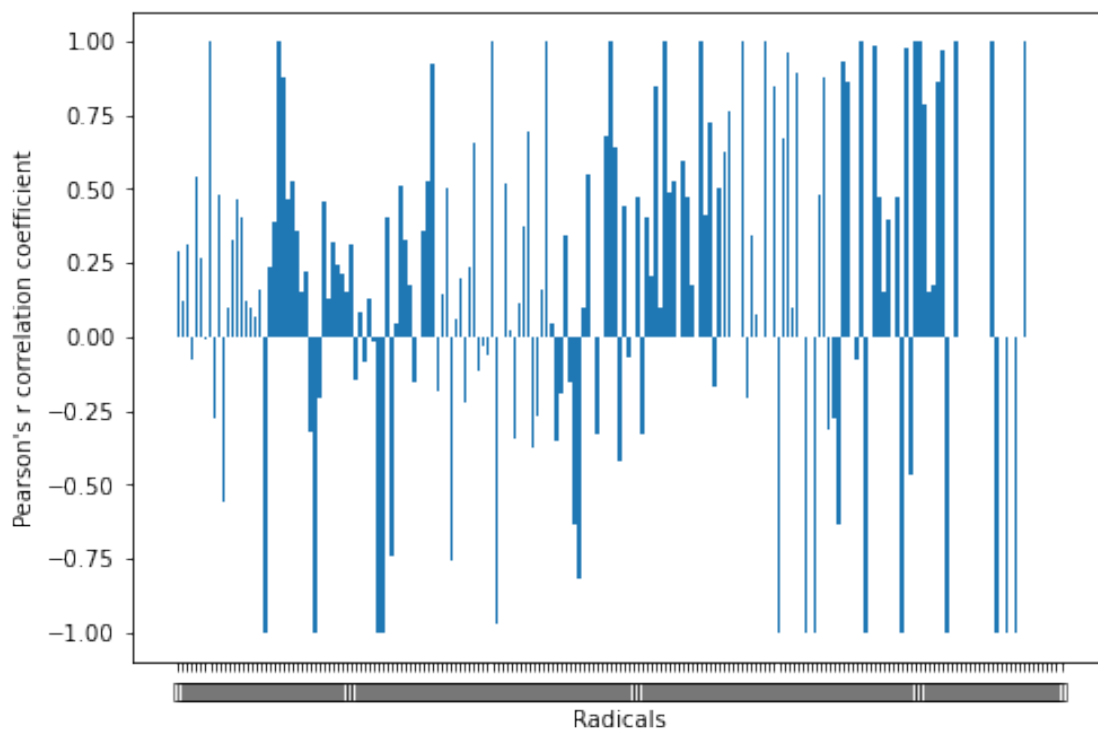
```

```
ave_p = sum(p_value_list)/len(p_value_list)
print("Average p value is: ", ave_p)
```

Average Pearson's r correlation coefficient is: 0.1542809522266968  
Average p value is: 0.3194279650468257

```
[156]: fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
ax.bar(radicals, r_value_list)
plt.ylabel("Pearson's r correlation coefficient")
plt.xlabel("Radicals")
```

```
[156]: Text(0.5, 0, 'Radicals')
```



```
[153]: # correlation analysis
from scipy.stats import pearsonr
stroke_cor = pearsonr(stroke_count, rt)
radical_cor = pearsonr(radical_removed, new_rt)

# study 1 correlation
print("Correlation for study 1 is: ", stroke_cor)

# study 2 correlation
```

```
print("Correlation for study 2 is: ", radical_cor)
```

Correlation for study 1 is: (0.3098711047884415, 8.966165864431881e-57)

Correlation for study 2 is: (0.2790367734354088, 7.186914042211598e-46)

```
[154]: # Check characters with more than 20 strokes
for i in char:
    if charlookup.getStrokeCount(i) >20:
        print(i)
```