

Q1) >>  $t = \text{table}$  (Name, Symbol, PeriodRow, GroupColumn, Type of Element, AtomicNumber, AtomicWeight - amu -, NaturalState, DISC  
over, electrons AddingToWhichEnergyLevel -, NaturalForming -, atomicRadius, pm -, electronegativity, x1stIonizationEnergy - kJ/mol, density - g/cm<sup>3</sup> -, meltingPoint - K -, boilingPoint - K -, specificHeat - J/Kg - K -);

>>  $\text{lme} = \text{fitlme}(t, \text{'YearDiscovered} \sim \text{electronsAddingToWhichEnergyLevel} +$

GroupColumn + Type of Element + (1 | NaturalState))

independent variables

↑  
random effect

Linear mixed-effects model fit by ML

Model information:

Number of observations	78
Fixed effects coefficients	5
Random effects coefficients	4
Covariance parameters	2

Formula:

YearDiscovered ~ 1 + GroupColumn + Type of Element + electronsAddingToWhichEnergyLevel + (1 | NaturalState)

Model fit statistics:

AIC	BIC	LogLikelihood	Deviance
834.76	851.26	-410.38	820.76

Fixed effects coefficients (95% CIs):

Name	{'(Intercept)'}
	{'GroupColumn'}
	{'Type of Element_Non-Metal'}
	{'Type of Element_Semimetal'}
	{'electronsAddingToWhichEnergyLevel_1'}

Estimate	1793.3
SE	32.688
tStat	54.861
DF	73

positive slope correlates every 1 unit inc.

in IV, ~ 'estimate' value inc in dependent var.  
there is NOT a statistically significant difference b/w group column and year discovered  
group column and year discovered

Random effects covariance parameters (95% CIs):

Group: NaturalState (4 Levels)  
Name1 Name2 Type

every 1 unit inc. in IV corresponds to a 41.07 decrease in year discovered

Estimate	41.07
Lower	-22.991
Upper	105.093

there is NOT a statistically significant difference b/w type of element (non-metals) and year discovered

Group: Error  
Name Estimate Lower Upper

{'Res Std'}	43.304	36.868	50.864
-------------	--------	--------	--------

there is A statistically significant difference b/w electrons adding to which energy level and year discovered

Estimate Column: → helpful in building equation

$$Y = B_0 + B_1 x_1 + B_2 x_2 + B_3 x_3 + B_4 x_4$$

↳ independent variable  
↳ dependent variable  
↳ stands for constant or intercept  
↳ represents the slopes

$$Y = 1793.3 + 0.87 x_1 - 41.07 x_2 + 13.93 x_3 + 20.1 x_4$$

$$Y = 1793.3 + 0.87(\text{GroupColumn}) - 41.07(\text{Type of Element - Non metal}) + 13.93(\text{Type of Element - semi}) + 20.12(\text{electrons adding to which energy level})$$

the ratio of the dependent variable over the respective independent variable

To answer this question, I used a mixed-effects model. I decided to use a regression model since I'm trying to determine if the independent variables in the question can predict the year that the element was discovered. Instead of just using a linear regression, I had to also consider the random effect of natural state using a random intercept model + the best way to do this is through a mixed-effects model. Based on the p-value column, for each independent variable (electrons, GroupColumn, and type of element), none of them can significantly predict the year discovered (due to a p-value less than the alpha value of 0.05). However, electronsAddingToWhichEnergyLevel can significantly predict the year discovered (due to the p-value greater than the alpha value of 0.05).

Electrons adding to which energy level can predict the year discovered – meaning that x electron adding to which energy level will positively correlate (as seen by the positive slope) with when the element was discovered. For example, aluminum and silicon have the same number of ‘electrons adding to which energy level?’ (3) and as we expect from our model, which positively correlates this relationship, the years that they were discovered are very similar to one another (1825 and 1823 respectively).

impliation of this

However, group column and type of element (independent variables) cannot predict the year discovered since their p-values are NOT statistically significant.

Explanation  
why  
I chose  
this test

Q 2)

$[b, \log_1, H, \text{stats}] = \text{Coxphfit}(\text{independent-variable (1 of electrons), dependent-variable (time until failure)})$

↓  
cox

$\gg [b, \log_1, H, \text{stats}] = \text{Coxphfit}(\text{electrons Adding To Which Energy Level - , Year Discovered})$

$\gg \text{ConfB: } 0.0058$

\*assuming that hazard is constant\*

$\beta_{\text{e}}: -0.6077$

→ Key value

$8e: 0.0764$

$z: -7.9577$

$P: 1.7521 \times 10^{-15}$  → this is the p-value, so there is a significant effect

Each increase(1) of the number of electrons for time until discovery is  $e^{-0.6077} = 0.54$  times more hazardous. The fewer the number of electrons, the earlier the element was discovered.

(Q3) → PCA

Must perform PCA b/c we want to take many dimensions + condense into only a couple so that we can make inferences about them

Principal Components are ordered on degrees of variation + eigenvalues

(PC1)

Differences along the 1<sup>st</sup> principal component axis are more disparate than differences along the second principal component axis (PC2)

After spending 6 hours on this question (\*insert crying emoji\*)...

```
% create a table with just the variables given in the question  
t2 = table(electronsAddingToWhichEnergyLevel_, atomicRadius_pm_, electronegativity, x1stIonizationEnergy_kJ_mol_, density_g_cm3_, meltingPoint_K_, boilingPoint_K_, specificHeat_J_kg_K_);  
  
% create an array from the table  
t2_mat=table2array(t2);  
↳ created a new table with only what we are interested in (all continuous, no categorical)  
  
% run the PCA  
[W,pc2,latent,tsquared,explained,mu] = pca(t2_mat);  
  
% PCA Code #1  
↳ PCA Code  
  
pc2clustersx = clusterdata(pc2(:,1:2),'maxclust',8,'linkage','av');  
labels2 = cellstr( num2str([1:118]' ));  
gscatter(pc2(:,1),pc2(:,2),pc2clustersx)  
text(pc2(:,1), pc2(:,2), labels2, 'VerticalAlignment','bottom', 'HorizontalAlignment','right')  
xlabel('First Principal Component');  
ylabel('Second Principal Component');  
title('Principal Component Scatter Plot Elements');
```

↳ need a numerical array  
↳ to graph  
↳ 118 elements

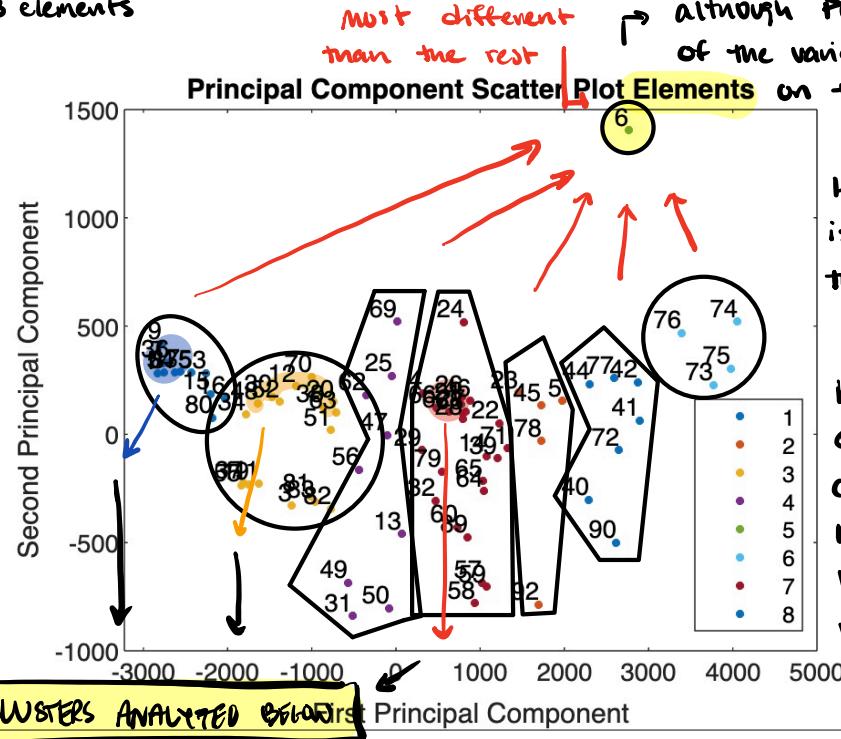
percentage of variation explained by PC1

explained =

94.7462
4.0024
1.2276
0.0233
0.0005
0.0001
0.0000
0.0000

percentage of variation explained by PC2

percentage of variation explained by PC3



CARBON IS V. UNIQUE

although PC1 accounts for 94% of the variation (as seen by magnitude on the x-axis), PC2 does account for 4%. (vertically: y-axis)  
↳ so it seems like carbon is different from the rest of the group.

Why? → Carbon's

ionization energy and electron affinity is high, can form 4 bonds, can link to itself, form polymers, has a high melting + boiling point. Has a valence of 4 → helps make different types of bonds

Senthilkumar, Siva

Blue: left - most <sup>cluster</sup>  
 Red: right - most <sup>between</sup> <sub>cluster</sub>  
 Yellow: middle cluster

**Periodic Table of the Elements**

©2014  
Todd Helmenstine  
scienotes.org

Based on PNA and Periodic Table: Characteristics

Tight Grouped Blue Clusters →

Nonmetal, similar + high ionization energies, 'Nof' involved in H-bonding, similar atomic radius, tend to be gases, similar number of valence electrons

Tight Grouped Red Clusters →

Metals, similar + lower ionization energies compared to blue group, typically 3-4 electrons added per row, many d and f group elements, similar atomic radius

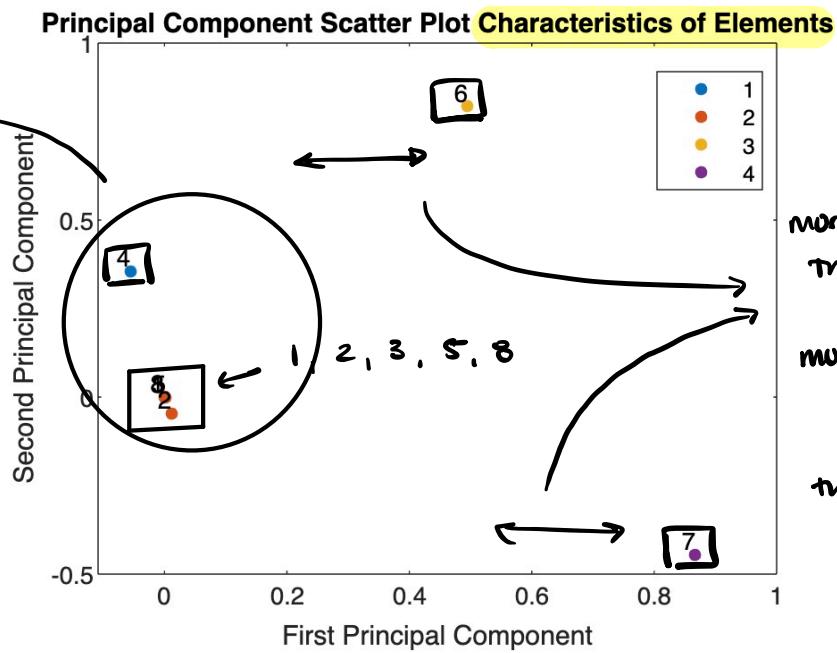
Tight Grouped Yellow Clusters →

Mix of metals + nonmetals, 3-5 electrons added, similar atomic radius, many have the same # of valence electrons (2)

%% Q3

```
% create a table with just the variables given in the question  
t2 = table(electronsAddingToWhichEnergyLevel_, atomicRadius_pm_, electronegativity, x1stIonizationEnergy_kJ_mol_, density_g_cm3_, meltingPoint_K_, boilingPoint_K_, specificHeat_J_kg_K_);  
  
% create an array from the table  
t2_mat=table2array(t2); → convert to array  
  
% run the PCA  
[W2,pc2,latent,tsquared,explained,mu] = pca(t2_mat);  
  
% PCA Code #2  
  
pcclusters = clusterdata(W2(:,1:2), 'maxclust', 4, 'linkage', 'av');  
labels = cellstr( num2str([1:8]) );  
gscatter(W2(:,1),W2(:,2),pcclusters); → 8 characteristics (as seen by column)  
text(W2(:,1), W2(:,2), labels, 'VerticalAlignment','bottom', 'HorizontalAlignment','right')  
xlabel('First Principal Component');  
ylabel('Second Principal Component');  
title('Principal Component Scatter Plot Characteristics of Elements');  
  
→ 4 clusters  
→ 4 columns
```

the columns in this circle are more similar to each other than one might assume b/c PC2 only accounts for 67% in variation



Most Similar: (1) Electrons, Atomic Radius, Electronegativity, 1st Ionization Energy, Density, Specific Heat  
(2) (3) (4)  
(5) (6)

→ makes sense that all of these categories represent factors intrinsic to the element itself (as in the # of electrons is v. similar to electronegativity and ionization energy) + density + specific heat are v. similar to the electron structure as well. The reason why melting and boiling point are so disparate from the other group is b/c these two group's characteristics are dependent on noncovalent interactions b/w molecules.

(Q 4)

linear regression w interaction

independent variables

Type of Element and # of Electrons

Electronegativity?

Multiple Independent Variables

dependent variable

~~All-way ANOVA~~ → All categorical variables, this data set does NOT have just Categorical variables

Linear Regression Model → Can't run logistic or mixed-effects models  
→ b/c no random effect or categorical dependent variables  
Yes!

>> lm = lm(lt, 'electronegativity ~ TypeOfElement \* electronsAddingToWhichEnergyLevel - 1);

interaction is when two independent variables are multiplied together

IV's alone

Interesting b/c type-of element has + slope (every 1 unit inc. in DV, .3, .42 or 0.27 inc. in electronegativity. However, every 1

Linear regression model:

electronegativity ~ 1 + TypeOfElement \* electronsAddingToWhichEnergyLevel

Estimated Coefficients:

Independent variables

Interaction term

	Estimate	SE	tStat	pValue
(Intercept)	1.6104	0.21758	7.4013	6.2101e-11
TypeOfElement_Non-Metal	1.4271	0.39763	3.5889	0.00053478
TypeOfElement_Semimetal	0.26684	0.61491	0.43395	0.66534
electronsAddingToWhichEnergyLevel	-0.026038	0.048177	-0.54047	0.59018
TypeOfElement_Non-Metal:electronsAddingToWhichEnergyLevel	-0.038393	0.11163	-0.34392	0.73169
TypeOfElement_Semimetal:electronsAddingToWhichEnergyLevel	0.072222	0.14141	0.51074	0.61076

Interaction terms

interesting that depending on the type of element (non or semi) it alters if the model's slope w interaction is pos. or neg.

Number of observations: 98, Error degrees of freedom: 92

Root Mean Squared Error: 0.438

R-squared: 0.546, Adjusted R-Squared: 0.521

F-statistic vs. constant model: 22.1, p-value = 1.75e-14

This means that electronegativity

can be predicted by non metal alone  
(no interaction)

From the output, it can be noted that NONE of the independent variables or different combinations of them (interactions), other than TypeOfElement\_Non-Metal can predict electronegativity.

↳ b/c the p-values are less than the  $\alpha$ -value of 0.05 (except for non-metal)

↳ example many nonmetals (N, O, F, Ne) have electroneg

in the 3.2 - 3.8 range.

>> anova (lm)

ans =

4x5 table

→ does not split into metals or nonmetals

TypeOfElement  
electronsAddingToWhichEnergyLevel\_  
TypeOfElement:electronsAddingToWhichEnergyLevel\_  
Error

	SumSq	DF	MeanSq	F	pValue
TypeOfElement	17.183	2	8.5914	44.771	2.6387e-14
electronsAddingToWhichEnergyLevel_	0.073275	1	0.073275	0.38184	0.53815
TypeOfElement:electronsAddingToWhichEnergyLevel_	0.084495	2	0.042247	0.22015	0.80282
Error	17.655	92	0.1919		

Le interaction cannot predict electronegativity

p-value less than α-value of 0.05 →  
cannot predict electronegativity

## FINAL MATLAB SCRIPT :

```
% Q1
clear ElementsInThePeriodicTable opts
t = table(Name, Symbol, PeriodRow, GroupColumn, TypeOfElement, AtomicNumber, AtomicWeight_amu_, NaturalState, YearDiscovered, Discoverer, electronsAddingToWhichEnergyLevel_, NaturalForming_, atomicRadius_pm_, ionizationEnergy_kJ_mol_, density_g_cm3_, meltingPoint_K_, boilingPoint_K_, specificHeat_J_kg_K_);
lm = fitlm(t, 'YearDiscovered~electronsAddingToWhichEnergyLevel_+GroupColumn+TypeOfElement+(1|NaturalState)');

% Q2
[b,logl,H,stats] = coxphfit(indepent_variable (number of electrons), dependent_variable (time until failure))
[b,logl,H,stats] = coxphfit(electronsAddingToWhichEnergyLevel_, YearDiscovered);

% Q3
% create a table with just the variables given in the question
t2 = table(electronsAddingToWhichEnergyLevel_, atomicRadius_pm_, electronegativity, x1stIonizationEnergy_kJ_mol_, density_g_cm3_, meltingPoint_K_, boilingPoint_K_, specificHeat_J_kg_K_);

% create an array from the table
t2_mat=table2array(t2);

% run the PCA
[W2,pc2,latent,tsquared,explained,mu] = pca(t2_mat);

% PCA Code #1
pcclustersx = clusterdata(pc2(:,1:2), 'maxclust', 8, 'linkage', 'av');
labels2 = cellstr( num2str([1:118]' ) );
gscatter(pc2(:,1),pc2(:,2),pcclustersx)
text(pc2(:,1), pc2(:,2), labels2, 'VerticalAlignment','bottom', 'HorizontalAlignment','right')
xlabel('First Principal Component');
ylabel('Second Principal Component');
title('Principal Component Scatter Plot Elements');

% PCA Code #2
pcclusters = clusterdata(W2(:,1:2), 'maxclust', 4, 'linkage', 'av');
labels = cellstr( num2str([1:8]' ) );
gscatter(W2(:,1),W2(:,2),pcclusters)
text(W2(:,1), W2(:,2), labels, 'VerticalAlignment','bottom', 'HorizontalAlignment','right')
xlabel('First Principal Component');
ylabel('Second Principal Component');
title('Principal Component Scatter Plot Characteristics of Elements');

% Q4
lm = fitlm(t, 'electronegativity~TypeOfElement*electronsAddingToWhichEnergyLevel_');
anova(lm)
```

