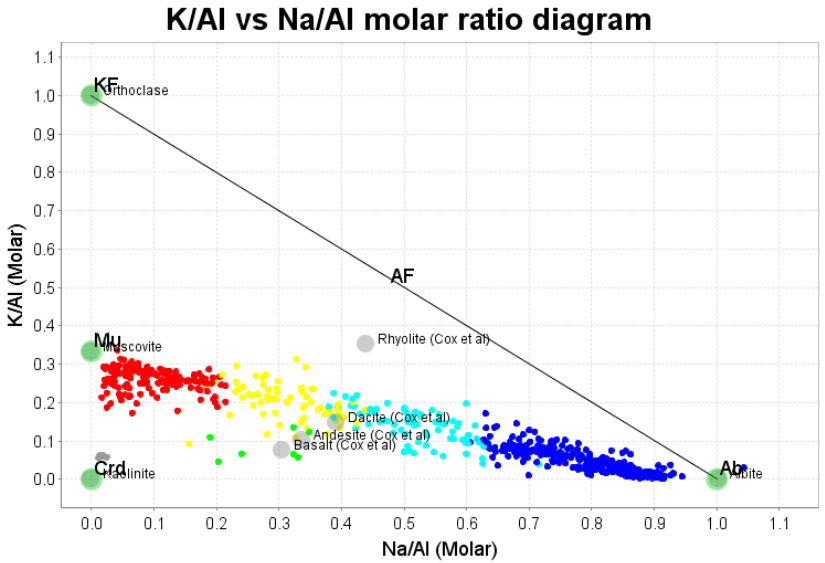
Mathematical Tools and Geochemistry

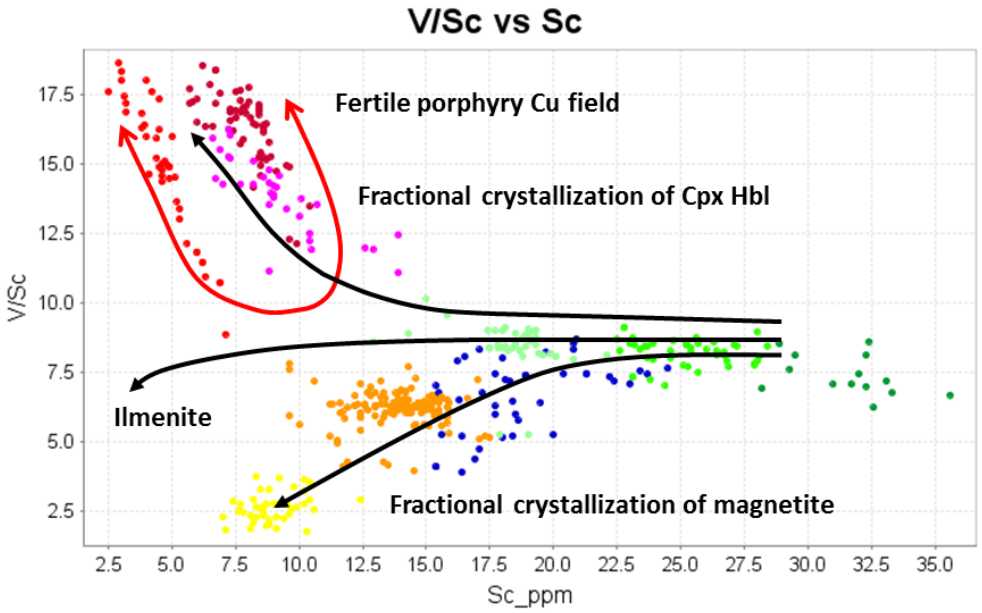
November 2022

Motivation

Rock Sample Media

- Geological Observation
  - Lithology
  - Alteration
  - Ore Characterization
- Geological Context
  - Regional and Tectonic Setting
  - Metallogenic Belts
- Geochemical Results
  - Rock Classification - Alteration
  - Lithogeochemistry
  - Pathfinders elements





Motivation

- Other Sample Media
  - Soils
  - Sediments
  - Biogeochemistry
  - Termites Nest
  - Others

- Biasing the sample
  - Sieving or other mineral separation
  - Strong and weak digestions

- Multielement Data Analysis

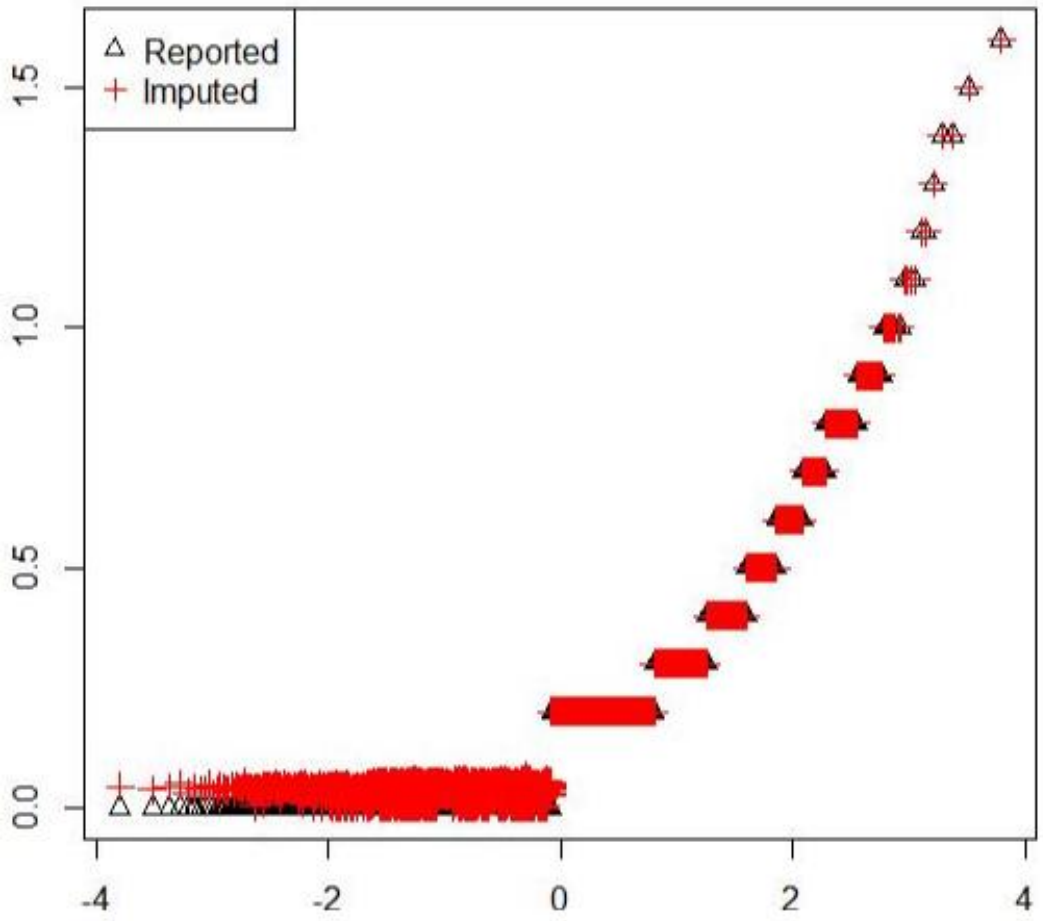
Mathematics

- Workflow
  - EDA
  - Imputation
  - Transformation
  - Principal Component Analysis
  - Machine Learning Technics
    - Supervised and Unsupervised
      - Regressions Models
      - Tree Classification
      - Random Forest Classification-Regression
      - Neural Networks
      - Clusters

Imputation and Transformation

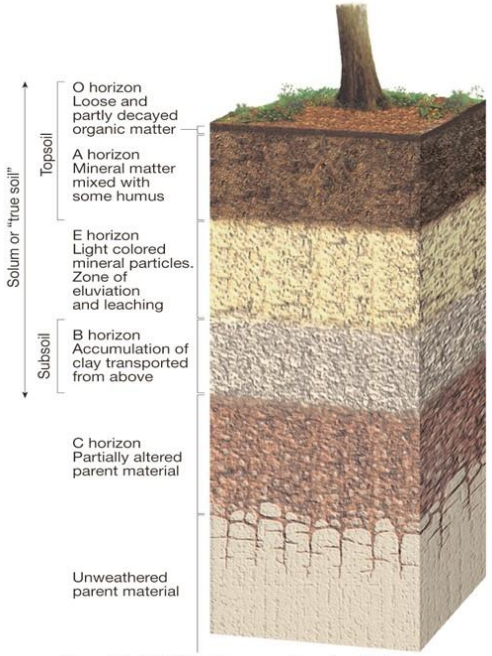
- Imputing below detection limit values.
  - ½ Detection Limit ?
    - No theorical base
    - Can lead to bias
    - Use of other methods
  - Distributional Methods
  - Robust Methods
    - Zcompositions R package

- Transformation
  - Normalization N(0,1)
  - Centered Log Ratio



Unsupervised Example

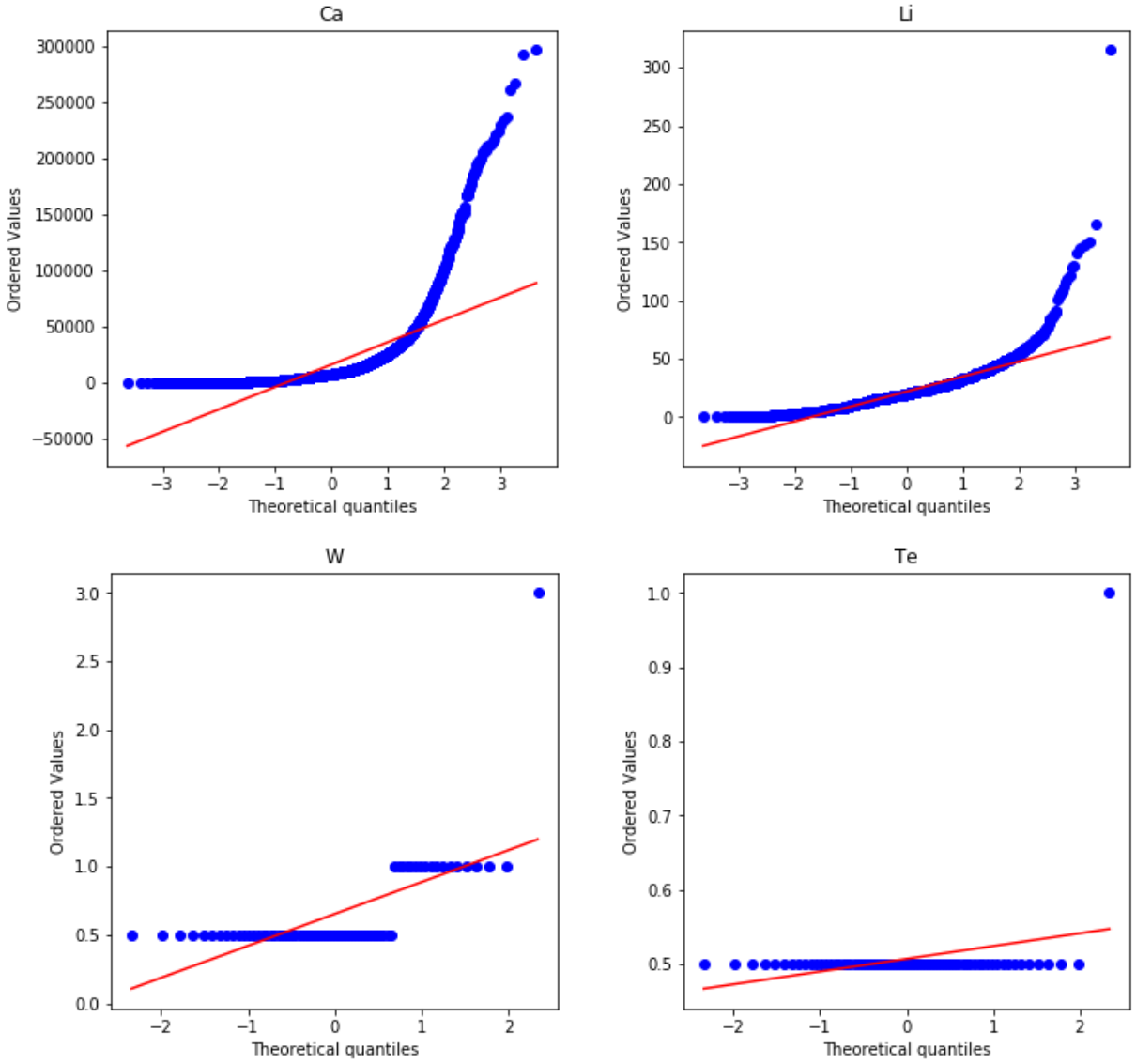
- Soil Samples
  - 4800 soil samples
  - Horizon A
  - Geochemistry, Aqua Regia and ICP-MS



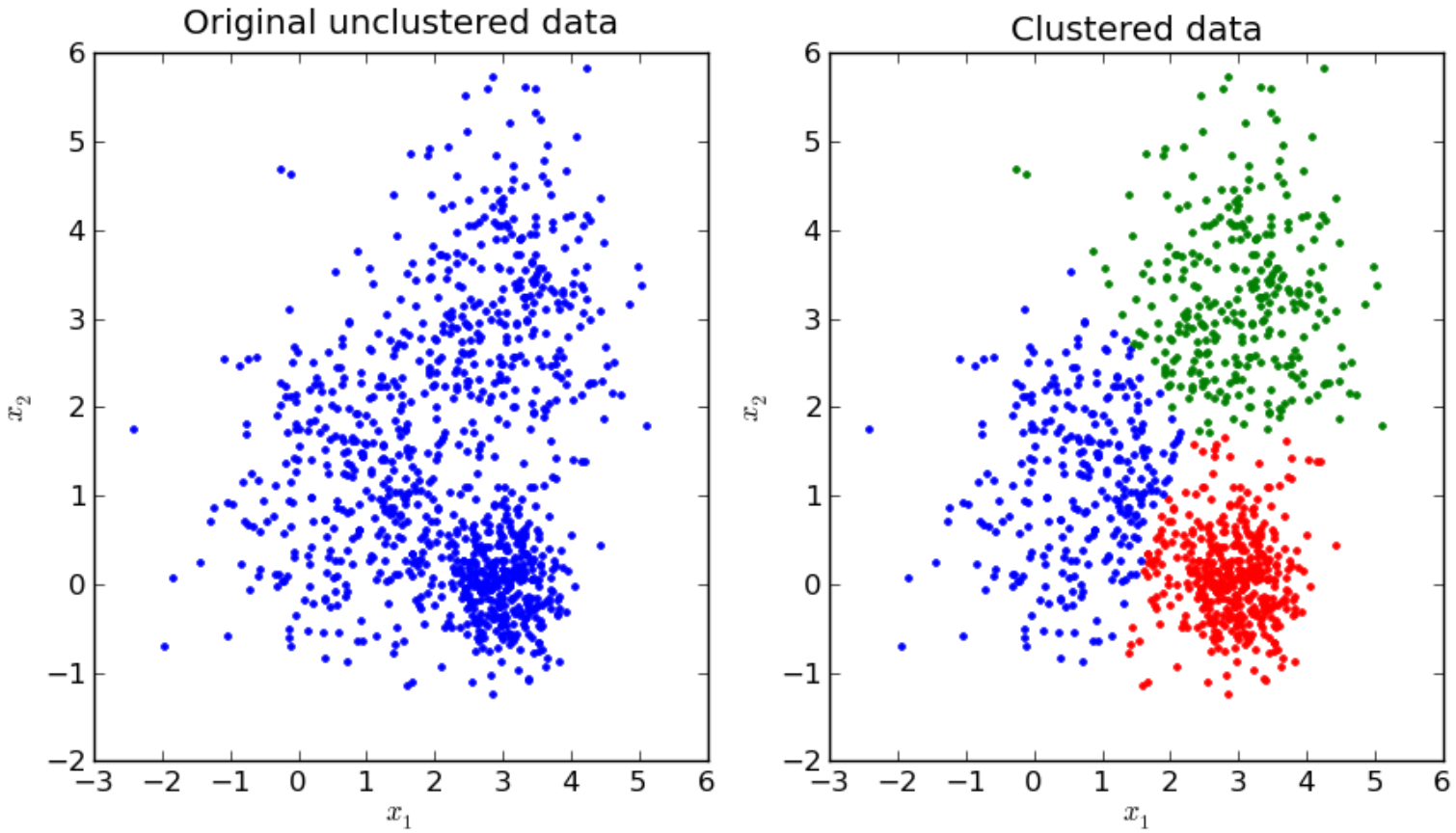
Latitude	Longitude	CollDate	LandCover1	LandCover2	Ag	Al	As	Ba	Be	Bi	C	Ca
31.3146	-87.1166	39862	Planted/Cultivated	Pasture/Hay	<1	8700	1.6	49	0.2	0.08	7100	700
33.8523	-86.9041	39870	Developed	Low Intensity Residential	<1	11500	1.5	118	0.4	0.1	8800	200
31.2179	-85.8788	39862	Planted/Cultivated	Fallow	<1	10500	1.2	34	0.2	0.07	6500	500
33.0992	-86.9976	39867	Forested Upland	Evergreen Forest	<1	41700	3.7	316	1.1	0.16	11400	500
34.5501	-87.6779	39869	Developed	Low Intensity Residential	<1	15100	2.7	179	0.6	0.09	13600	300
34.3577	-87.313	39870	Forested Upland	Deciduous Forest	<1	27800	4.3	160	0.5	0.19	17100	200
31.0634	-85.6928	39863	Planted/Cultivated	Fallow	<1	21400	2.7	50	0.3	0.16	8900	400
34.2848	-85.6099	39868	Forested Upland	Mixed Forest	<1	18300	3.7	185	0.5	0.21	45900	2400
32.7952	-86.9501	39867	Forested Upland	Evergreen Forest	<1	17800	2.9	77	0.4	0.15	23200	<100
31.7278	-87.1307	39864	Developed	Low Intensity Residential	<1	12300	3	41	0.2	0.12	8900	400
30.4554	-87.5496	39861	Developed	Low Intensity Residential	<1	14800	2.3	54	0.2	0.12	31000	400
32.1407	-85.5465	39863	Planted/Cultivated	Fallow	<1	9100	1.5	43	0.2	0.08	6000	<100
32.9905	-86.651	39867	Forested Upland	Mixed Forest	<1	73300	23.9	412	2.4	0.31	29900	500
32.5952	-85.5658	39864	Planted/Cultivated	Pasture/Hay	<1	14500	2.3	88	0.4	0.12	36600	3200
32.392	-87.4686	39865	Developed	Low Intensity Residential	<1	61800	10	213	1.6	0.4	22300	2500
34.0736	-87.9234	39869	Forested Upland	Deciduous Forest	<1	15300	1.9	171	0.6	0.08	12500	200
34.9852	-87.5194	39869	Planted/Cultivated	Pasture/Hay	<1	13600	2.9	58	0.2	0.06	7900	400
31.3746	-88.0048	39861	Developed	Low Intensity Residential	<1	12300	1.5	157	0.4	0.09	19500	200
33.3123	-87.8587	39865	Planted/Cultivated	Pasture/Hay	<1	9800	1.6	126	0.4	0.08	14700	700
33.6629	-85.66	39868	Planted/Cultivated	Fallow	<1	58200	10.6	384	1.8	0.27	19300	1100
30.8919	-88.0869	39861	Developed	Low Intensity Residential	<1	2600	<0.6	22	<0.1	<0.04	5400	<100
32.6467	-86.5986	39867	Planted/Cultivated	Pasture/Hay	<1	9300	1.5	80	0.3	0.09	8500	200
33.0284	-85.6592	39867	Developed	Low Intensity Residential	<1	51600	2.6	182	0.9	0.19	10100	<100
31.9654	-87.0744	39864	Planted/Cultivated	Pasture/Hay	<1	69000	5.9	439	1.9	0.38	29000	73300
34.5247	-86.4085	39868	Planted/Cultivated	Pasture/Hay	<1	21500	4.5	110	0.4	0.2	6700	1100
31.0277	-87.061	39862	Developed	Low Intensity Residential	<1	10700	1.5	50	0.1	0.09	7100	<100
34.0374	-87.2118	39870	Forested Upland	Evergreen Forest	<1	18000	2.1	121	0.4	0.09	9400	200
31.4932	-86.1309	39862	Planted/Cultivated	Pasture/Hay	<1	15400	2.6	131	0.7	0.1	2200	400
33.491	-87.5339	39870	Planted/Cultivated	Fallow	<1	27100	4.8	201	0.7	0.14	16000	600
32.8294	-86.1011	39867	Developed	Low Intensity Residential	<1	88900	2.8	282	2.2	0.14	11900	900
34.3422	-87.9365	39869	Developed	Low Intensity Residential	<1	15900	2.8	176	0.5	0.12	7600	300
34.3654	-87.0215	39870	Planted/Cultivated	Pasture/Hay	<1	39600	10.2	282	1.1	0.22	7900	600

Exploratory Data Analysis

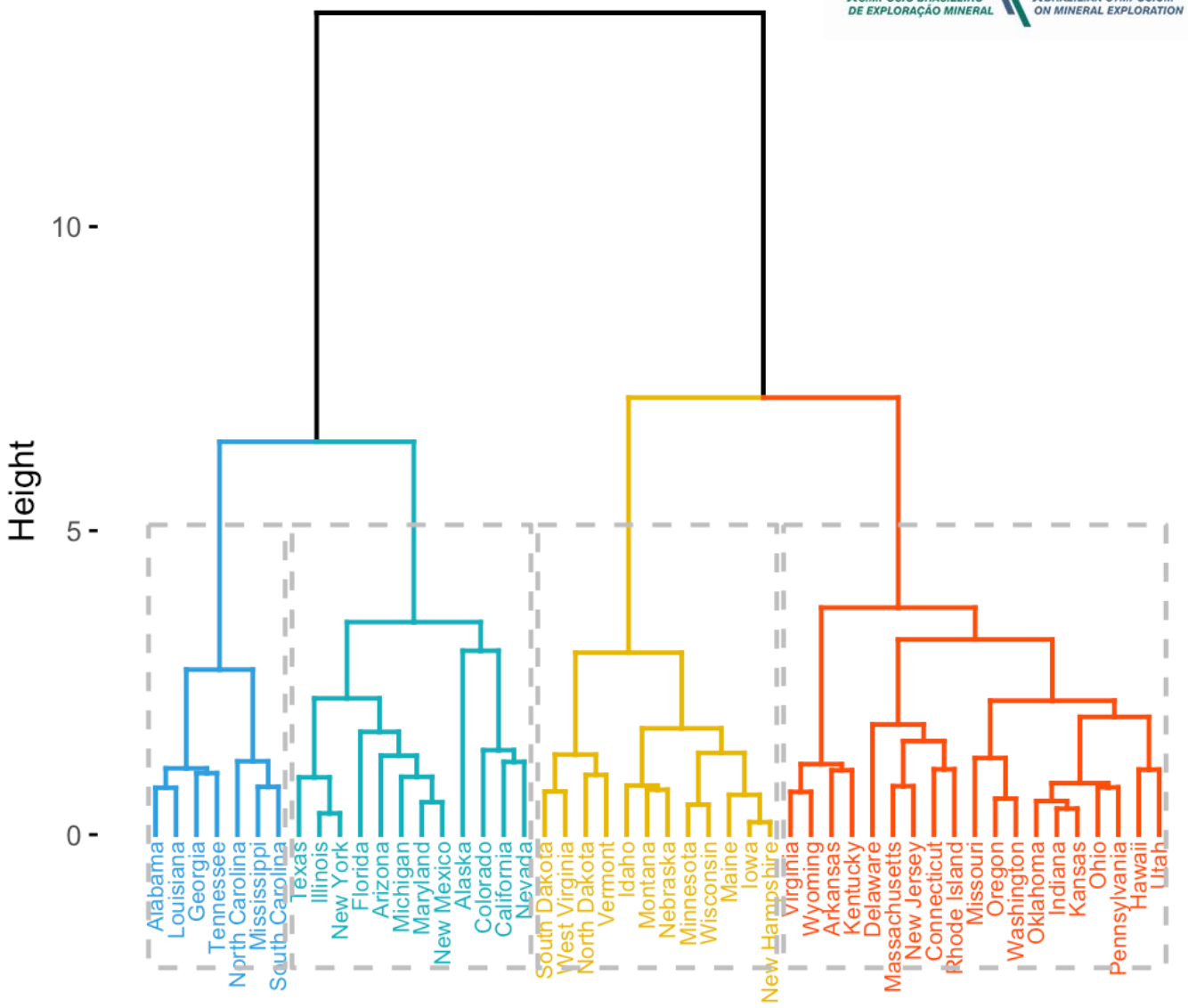
- Useless Analytes
  - Due to the geochemical method not all analytes has good distribution.
- QQ plots
  - Distribution of results.
  - 8 analytes were censored:
    - Bi, U, In, W, Re, Sn, Ta, Te



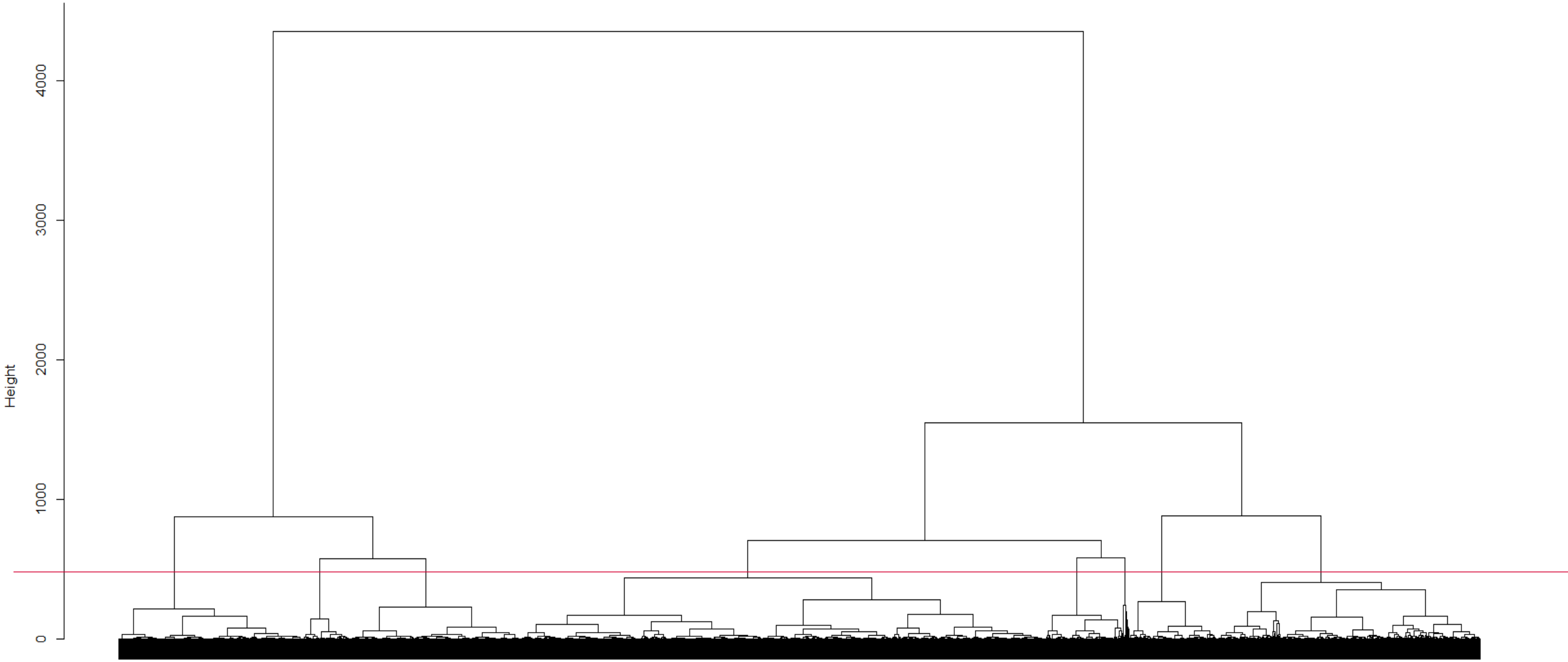
Clustering



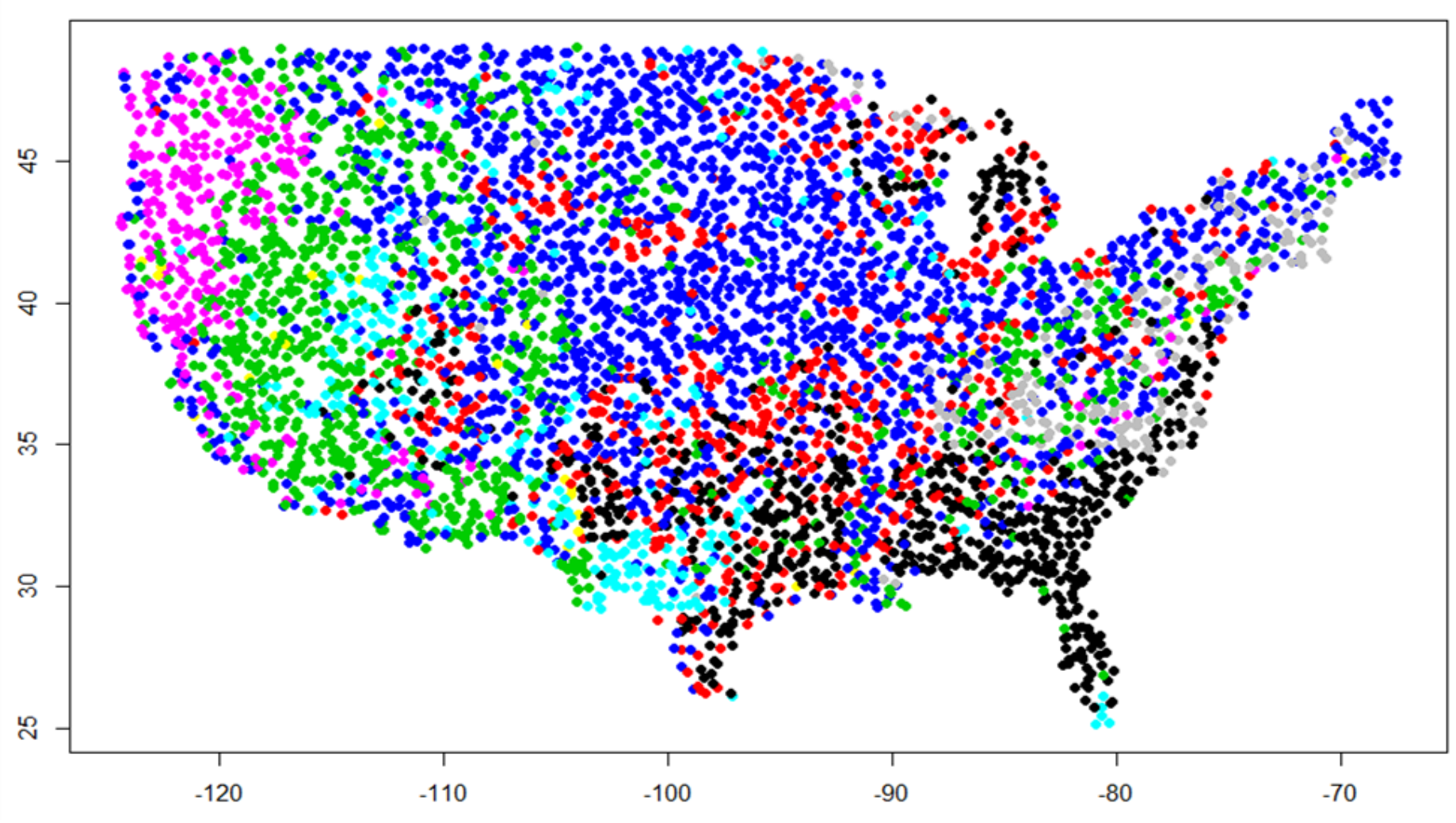
Clustering



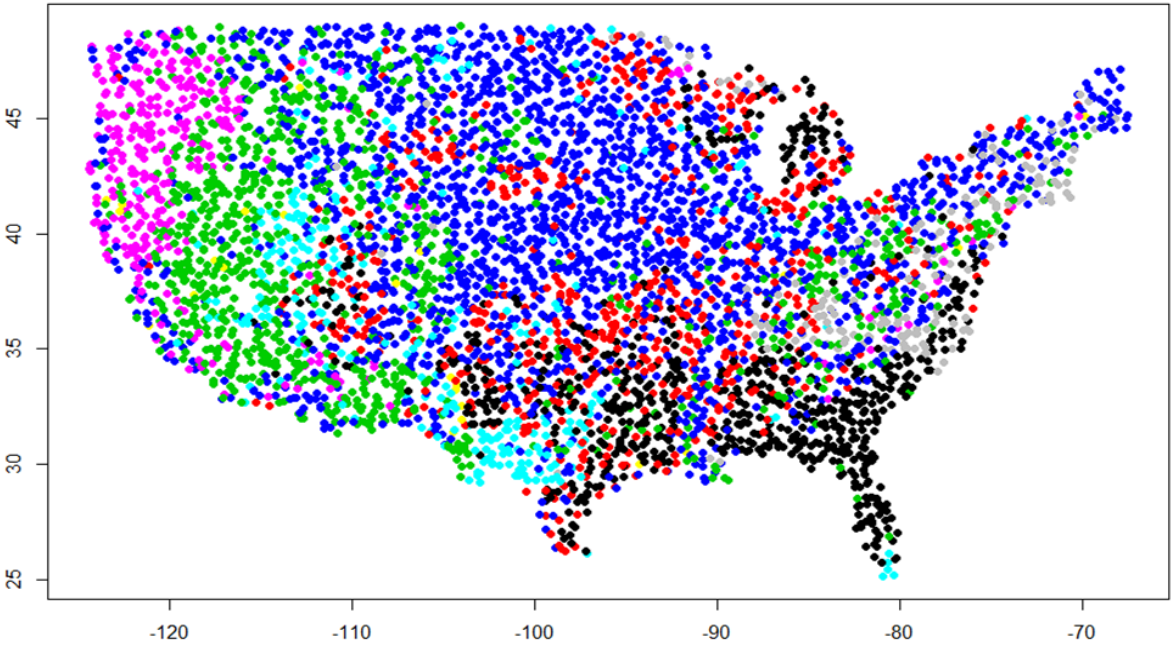
Cluster Dendrogram

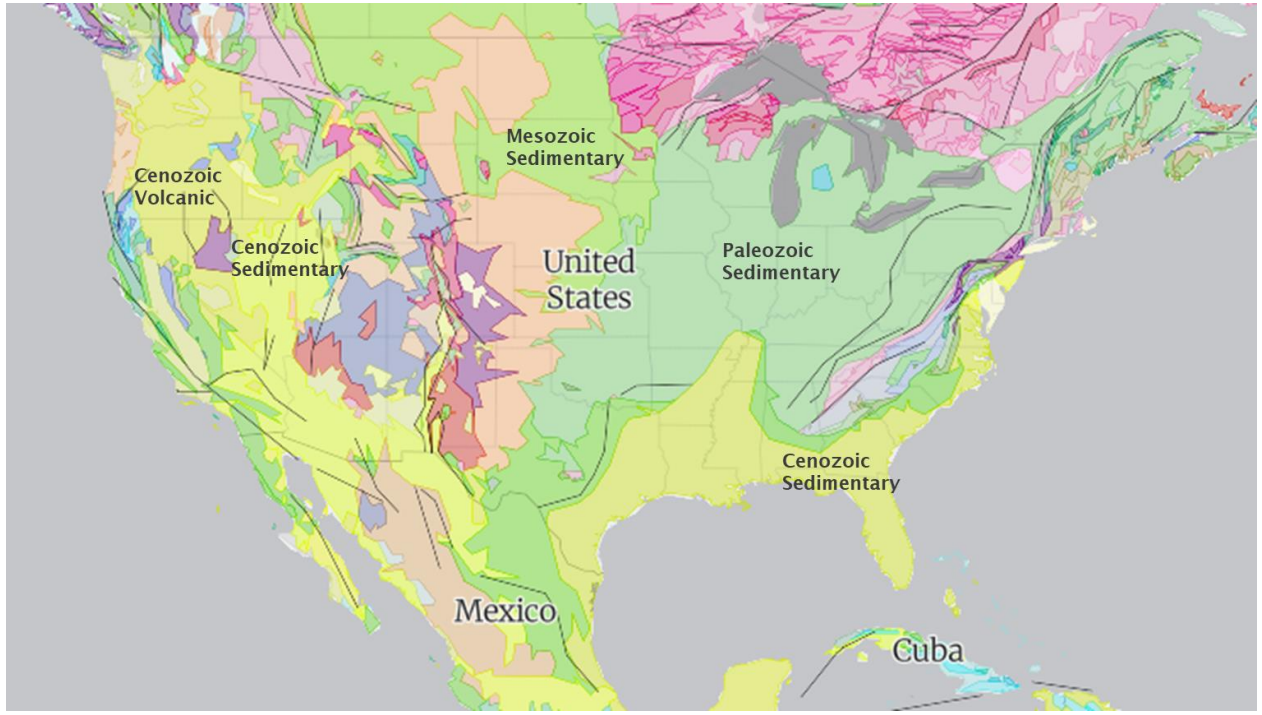


Relationship between soil and rocks under



Relationship between soil and rocks under





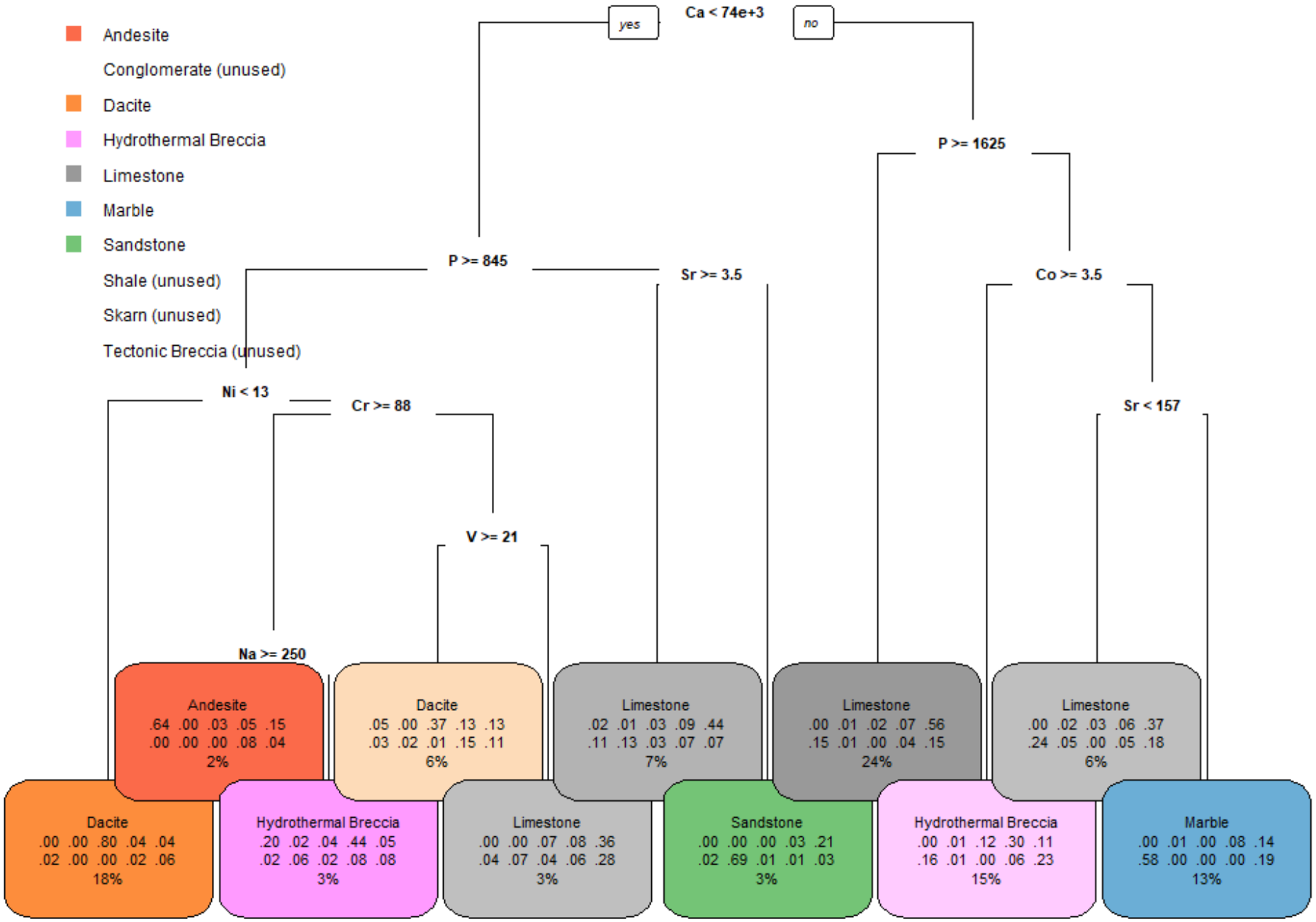
Supervised Example - Decision Tree Models

- Lithology prediction.
- Splitting data:
  - Training Data 70%
  - Testing Data 30%
- Splitting by target attribute.
- Aqua regia digest and ICP-MS

Litho	Ag	Al	As	B	Ba	Be	Bi	Ca	Cd	Ce	Co	Cr
Hydrothermal Breccia	6.51	11200	958	90	30	1.3	2	54400	21.1	40	30	104
Hydrothermal Breccia	9.60	11200	1120	50	20	2.4	1	61000	24.2	60	39	164
Limestone	0.41	6600	158	20	40	0.6	1	175000	0.6	10	9	33
Limestone	0.41	5400	160	10	10	0.5	1	148000	0.25	20	14	67
Limestone	2.40	4200	134	120	20	0.6	2	184000	6.2	10	5	22
Limestone	2.81	1700	166	20	20	0.5	1	180000	2	10	4	12
Limestone	0.10	3400	41	10	30	0.25	2	188000	0.25	10	5	17
Limestone	0.10	4300	116	10	20	0.25	3	212000	0.25	10	5	14
Limestone	7.99	1300	172	5	20	0.5	3	163000	6.5	10	4	7
Limestone	0.10	2100	100	10	40	0.25	1	203000	0.7	10	4	11
Limestone	0.31	7400	246	10	10	0.5	3	194000	0.25	20	13	42
Limestone	0.99	2000	169	10	10	0.25	3	187000	2.4	10	4	7
Limestone	30.79	3400	764	10	10	1.1	6	75800	18.5	20	13	11
Limestone	5.79	2100	557	10	20	1.3	3	126000	3	10	8	23
Tectonic Breccia	6.51	3800	367	10	5	1.1	1	94000	60.6	10	3	10
Limestone	27.29	1100	1530	10	10	1.2	4	75700	190.5	10	2	9
Limestone	43.37	1500	4300	10	10	1.3	4	81000	87.9	20	3	16
Dacite	6.89	11100	2060	10	60	1.5	2	42300	5	30	12	27
Limestone	1.10	26100	825	5	5	0.8	1	109000	2.4	10	12	18
Limestone	1.30	20300	613	5	5	0.7	2	142000	0.6	10	15	14
Limestone	2.40	18900	593	5	5	0.5	3	121000	1.6	10	15	15
Limestone	50.06	2700	573	10	10	0.25	133	235000	218	10	0.5	20
Limestone	0.79	21900	135	5	5	0.5	1	113000	0.25	20	3	14
Limestone	1.61	13800	286	10	10	0.25	1	135000	1.6	20	3	12
Limestone	27.70	6000	1960	10	5	0.25	1	95000	128.5	10	0.5	8
Tectonic Breccia	5.79	7200	355	10	20	0.25	1	114000	23.4	10	2	8
Limestone	0.31	19300	681	5	5	0.5	1	155000	0.6	10	3	17

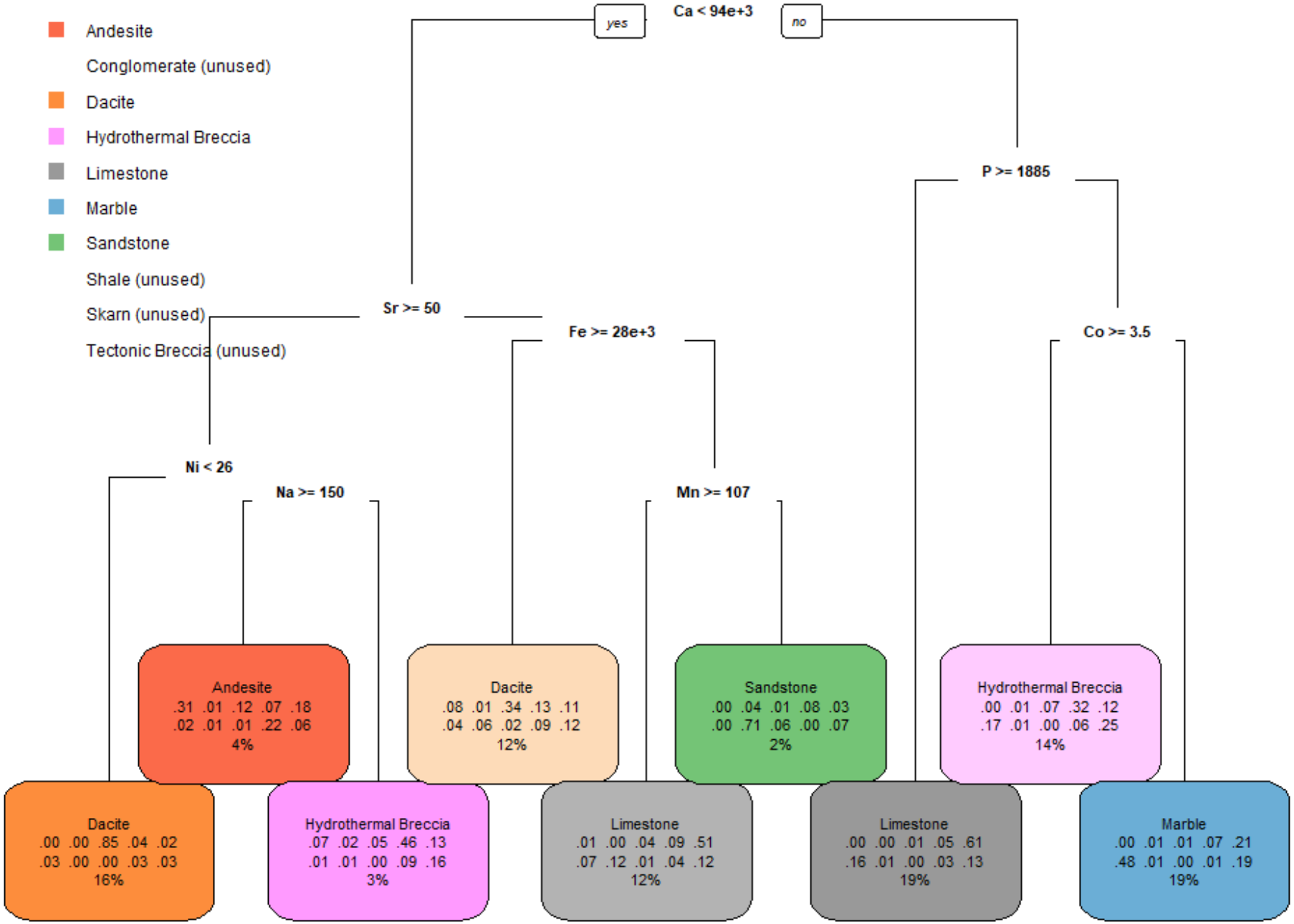
Decision Tree Models

- Lithology prediction:
- 53% over the test group.
- In four acids is over 90%.
- Different Splitting process results in different trees.



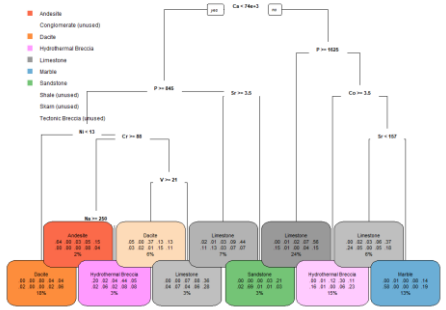
Decision Tree Models

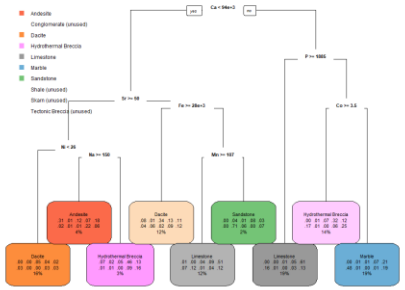
- Lithology prediction:
- Another tree: 52% in test group.
- Why not to use several trees?

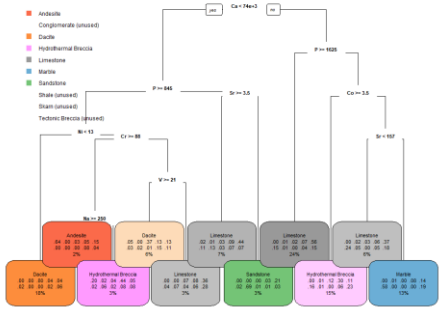


Random Forest

- Hundreds of trees from splitting train data randomly.
- Every tree has their own “weight” according to test results.







	Andesite	Conglomerate	Dacite	Hydrothermal Breccia	Limestone	Marble	Sandstone	Shale	Skarn	Tectonic Breccia
Andesite	68	0	4	13	4	0	3	1	5	2
Conglomerate	0	9	0	7	2	0	4	0	0	3
Dacite	0	0	712	18	7	4	1	0	3	13
Hydrothermal Breccia	6	0	27	289	31	32	7	0	2	43
Limestone	13	0	14	5	842	35	5	0	7	50
Marble	0	0	9	21	82	458	0	0	10	39
Sandstone	1	0	0	3	22	3	117	0	0	4
Shale	0	0	0	0	5	0	6	8	0	0
Skarn	3	0	22	14	10	8	0	0	109	6
Tectonic Breccia	2	2	32	57	91	74	5	0	9	250

Random Forest

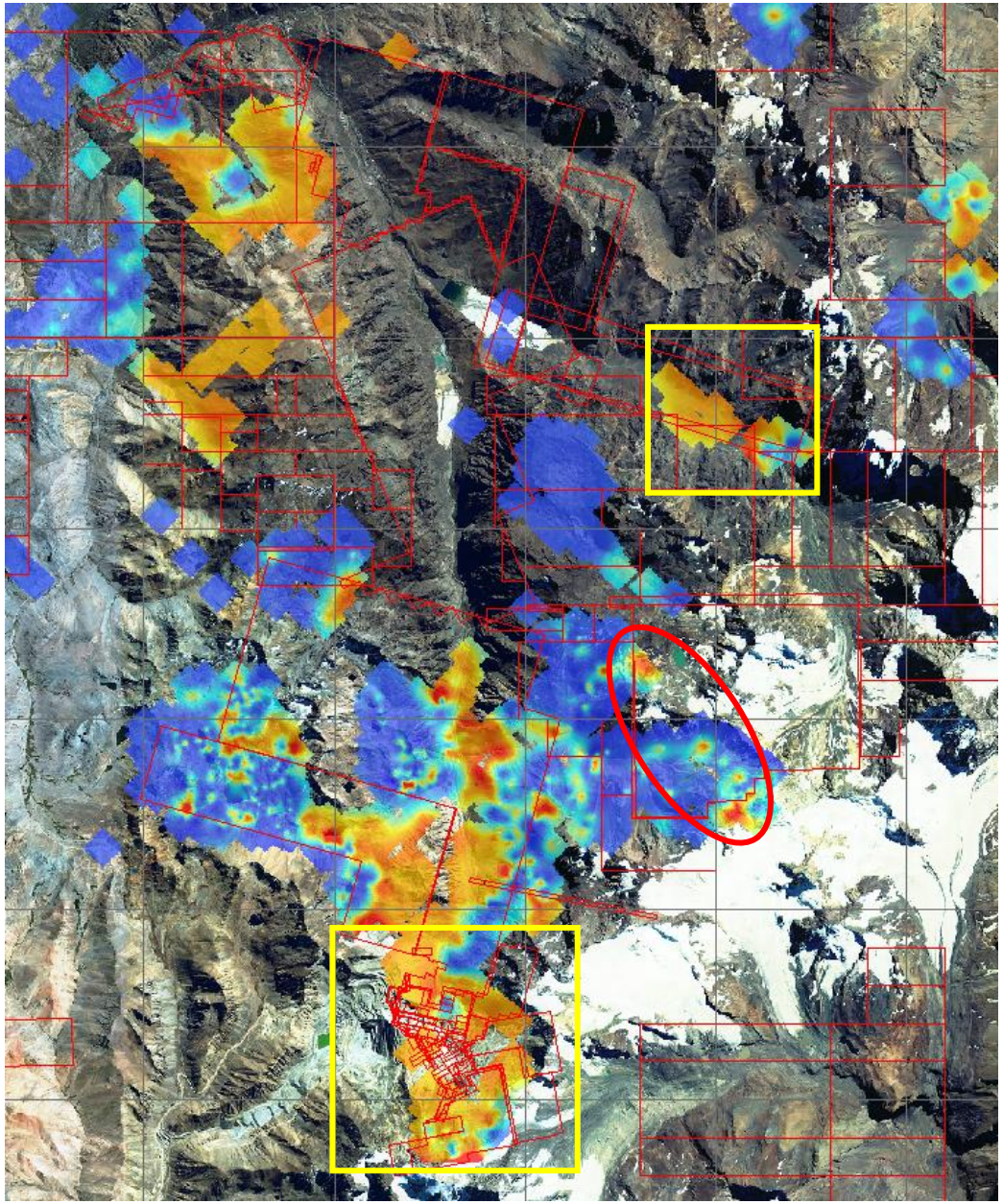
- Over 75% of data well assigned in test data.
- Near - Black box

	Andesite	Conglomerate	Dacite	Hydrothermal Breccia	Limestone	Marble	Sandstone	Shale	Skarn	Tectonic Breccia
Andesite	**68**	0	4	13	4	0	3	1	5	2
Conglomerate	0	**9**	0	7	2	0	4	0	0	3
Dacite	0	0	**712**	18	7	4	1	0	3	13
Hydrothermal Breccia	6	0	27	**289**	31	32	7	0	2	43
Limestone	13	0	14	5	**842**	35	5	0	7	50
Marble	0	0	9	21	82	**458**	0	0	10	39
Sandstone	1	0	0	3	22	3	**117**	0	0	4
Shale	0	0	0	0	5	0	6	**8**	0	0
Skarn	3	0	22	14	10	8	0	0	**109**	6
Tectonic Breccia	2	2	32	57	91	74	5	0	9	**250**

Cautionary Tales

- “Blind” Sample Media, apples and pears.
- Collecting comparable samples with comparable assays.

- Example, prediction models using different analytical methods:
  - ICP-MS and ICP-AES
  - Bi example



Cautionary Tales

- Predicting, wolfs and dogs.
- About geological context and geochemical signals.

Predicted: Wolf
True: Wolf

Predicted: Husky
True: Husky

Predicted: Husky
True: Husky

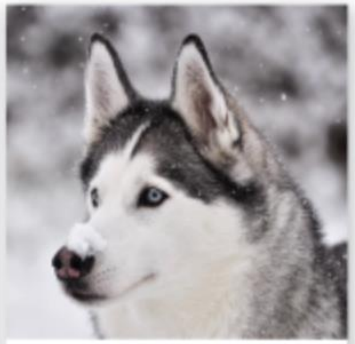
Predicted: Wolf
True: Wolf

Predicted: Wolf
True: Wolf

Predicted: Wolf
True: Wolf

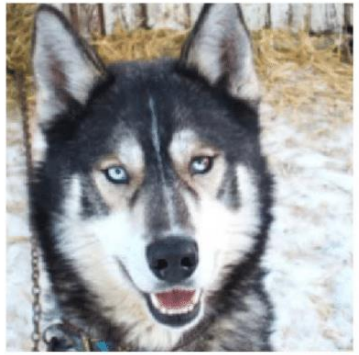
Predicted: Husky
True: Wolf

Predicted: Wolf
True: Wolf



Predicted: Wolf
True: Husky

Predicted: Husky
True: Husky



Thanks!