

Data Wrangling I

John DeForest

July 15, 2024

```
g1 = read.csv('C:\\Users\\John DeForest\\Desktop\\qbs103 R DS\\QBS103_GSE157103_genes.csv')
sm1 = read.csv('C:\\Users\\John DeForest\\Desktop\\qbs103 R DS\\QBS103_GSE157103_series_matrix-1.csv')
#gene, metadata

print(head(g1))
```

```
##      X COVID_01_39y_male_NonICU COVID_02_63y_male_NonICU
## 1    A1BG                      0.49                    0.29
## 2    A1CF                      0.00                    0.00
## 3    A2M                       0.21                    0.14
## 4    A2ML1                     0.04                    0.00
## 5 A3GALT2                      0.07                    0.00
## 6 A4GALT                       0.00                    0.00
## COVID_03_33y_male_NonICU COVID_04_49y_male_NonICU COVID_05_49y_male_NonICU
## 1                      0.26                      0.45                      0.17
## 2                      0.00                      0.01                      0.00
## 3                      0.03                      0.09                      0.00
## 4                      0.02                      0.07                      0.05
## 5                      0.00                      0.00                      0.07
## 6                      0.00                      0.00                      0.00
## COVID_06_.y_male_NonICU COVID_07_38y_female_NonICU COVID_08_78y_male_ICU
## 1                      0.21                      0.49                      0.12
## 2                      0.00                      0.01                      0.00
## 3                      0.08                      0.23                      0.08
## 4                      0.04                      0.03                      0.01
## 5                      0.00                      0.07                      0.00
## 6                      0.00                      0.00                      0.00
## COVID_09_64y_female_ICU COVID_10_62y_male_ICU COVID_11_52y_female_NonICU
## 1                      0.51                      0.10                      0.38
## 2                      0.01                      0.00                      0.02
## 3                      0.88                      0.13                      0.47
## 4                      0.02                      0.01                      0.03
## 5                      0.79                      0.15                      0.08
## 6                      0.00                      0.00                      0.00
## COVID_12_50y_male_ICU COVID_13_37y_male_NonICU COVID_14_55y_male_ICU
## 1                      0.45                      0.18                      0.23
## 2                      0.00                      0.00                      0.00
## 3                      0.16                      0.07                      0.22
## 4                      0.00                      0.01                      0.04
## 5                      1.75                      0.00                      0.93
## 6                      0.00                      0.00                      0.00
```

##	COVID_15_68y_male_ICU	COVID_16_48y_male_NonICU	COVID_17_54y_male_NonICU
## 1	0.42	0.41	0.63
## 2	0.00	0.01	0.02
## 3	0.07	0.58	0.15
## 4	0.00	0.00	0.02
## 5	0.15	0.19	0.00
## 6	0.03	0.00	0.00
##	COVID_18_70y_female_NonICU	COVID_19_51y_male_NonICU	COVID_20_62y_male_ICU
## 1	0.47	0.33	0.32
## 2	0.00	0.02	0.00
## 3	0.30	0.11	0.07
## 4	0.02	0.02	0.00
## 5	0.06	0.00	0.22
## 6	0.03	0.00	0.00
##	COVID_21_66y_male_ICU	COVID_22_43y_male_ICU	COVID_23_76y_male_ICU
## 1	0.18	0.09	0.18
## 2	0.00	0.00	0.01
## 3	0.00	0.06	0.03
## 4	0.00	0.00	0.00
## 5	0.37	0.06	0.07
## 6	0.03	0.00	0.03
##	COVID_24_55y_male_ICU	COVID_25_55y_male_ICU	COVID_26_41y_female_ICU
## 1	0.22	0.29	0.42
## 2	0.01	0.00	0.00
## 3	0.11	0.09	0.18
## 4	0.02	0.03	0.00
## 5	0.15	0.00	0.87
## 6	0.00	0.00	0.00
##	COVID_27_71y_female_ICU	COVID_28_63y_male_ICU	COVID_29_63y_female_ICU
## 1	0.16	0.18	0.35
## 2	0.01	0.00	0.00
## 3	0.23	0.18	0.03
## 4	0.01	0.05	0.03
## 5	0.18	0.45	0.15
## 6	0.00	0.00	0.03
##	COVID_30_54y_male_ICU	COVID_31_50y_male_ICU	COVID_32_72y_male_ICU
## 1	0.23	0.15	0.34
## 2	0.00	0.00	0.01
## 3	0.11	0.47	0.04
## 4	0.01	0.00	0.00
## 5	0.00	0.00	0.29
## 6	0.00	0.03	0.00
##	COVID_33_81y_male_NonICU	COVID_34_64y_female_NonICU	
## 1	0.35	0.36	
## 2	0.00	0.00	
## 3	0.30	0.11	
## 4	0.06	0.00	
## 5	0.26	0.12	
## 6	0.00	0.00	
##	COVID_35_58y_female_NonICU	COVID_36_68y_male_NonICU	COVID_37_87y_male_NonICU
## 1	0.26	0.18	0.20
## 2	0.00	0.01	0.00
## 3	0.51	0.09	0.09
## 4	0.02	0.00	0.07

## 5	0.16	0.08	0.31
## 6	0.00	0.00	0.00
## COVID_38_68y_male_ICU	COVID_39_80y_female_ICU	COVID_40_66y_male_ICU	
## 1	0.29	0.19	0.22
## 2	0.00	0.00	0.00
## 3	0.10	0.27	0.17
## 4	0.02	0.00	0.00
## 5	0.35	0.00	0.08
## 6	0.00	0.07	0.00
## COVID_41_74y_male_ICU	COVID_42_21y_female_ICU	COVID_43_83y_female_ICU	
## 1	0.19	0.24	0.29
## 2	0.00	0.01	0.00
## 3	0.14	0.33	0.00
## 4	0.00	0.01	0.00
## 5	0.19	0.39	0.11
## 6	0.00	0.00	0.00
## COVID_44_46y_male_ICU	COVID_45_62y_female_ICU	COVID_46_62y_male_ICU	
## 1	0.22	0.14	0.53
## 2	0.00	0.00	0.01
## 3	0.14	0.15	0.10
## 4	0.00	0.03	0.00
## 5	0.00	0.19	0.06
## 6	0.04	0.00	0.00
## COVID_47_78y_male_ICU	COVID_48_72y_female_ICU	COVID_49_73y_male_ICU	
## 1	0.08	0.19	0.48
## 2	0.01	0.00	0.00
## 3	0.04	0.06	0.09
## 4	0.03	0.01	0.03
## 5	0.60	0.23	0.00
## 6	0.00	0.06	0.00
## COVID_50_37y_male_ICU	COVID_51_58y_female_NonICU	COVID_52_71y_male_NonICU	
## 1	0.08	0.21	0.25
## 2	0.00	0.00	0.01
## 3	0.01	0.13	0.00
## 4	0.00	0.00	0.03
## 5	0.00	0.00	0.00
## 6	0.72	0.00	0.00
## COVID_53_35y_female_NonICU	COVID_55_62y_female_ICU	COVID_56_33y_female_NonICU	
## 1	0.25	0.09	0.28
## 2	0.00	0.00	0.00
## 3	0.64	0.09	0.16
## 4	0.10	0.01	0.09
## 5	0.00	0.00	0.23
## 6	0.00	0.00	0.00
## COVID_57_30y_female_NonICU	COVID_58_62y_male_NonICU	COVID_59_55y_male_NonICU	
## 1	0.42	0.39	0.33
## 2	0.00	0.00	0.00
## 3	0.27	0.08	0.10
## 4	0.01	0.00	0.00
## 5	0.19	0.00	0.07
## 6	0.05	0.00	0.00
## COVID_60_49y_male_NonICU	COVID_61_54y_female_NonICU	COVID_62_78y_female_ICU	
## 1	0.22	0.25	0.21
## 2	0.00	0.00	0.00

## 3	0.14	0.10	0.04
## 4	0.00	0.03	0.00
## 5	0.00	0.13	0.05
## 6	0.02	0.00	0.00
## COVID_63_39y_female_ICU COVID_64_65y_male_ICU COVID_65_84y_male_NonICU			
## 1	0.29	0.38	0.40
## 2	0.00	0.01	0.01
## 3	0.01	0.04	0.07
## 4	0.00	0.02	0.00
## 5	0.14	0.56	0.58
## 6	0.00	0.00	0.00
## COVID_66_66y_female_NonICU COVID_67_57y_male_ICU COVID_68_79y_male_ICU			
## 1	0.64	0.37	0.58
## 2	0.00	0.00	0.00
## 3	0.00	0.35	0.15
## 4	0.00	0.00	0.01
## 5	0.00	0.00	0.00
## 6	0.00	0.00	0.05
## COVID_69_77y_female_NonICU COVID_70_81y_male_NonICU COVID_71_37y_male_ICU			
## 1	0.52	0.27	0.07
## 2	0.00	0.00	0.01
## 3	0.29	0.07	0.12
## 4	0.02	0.00	0.01
## 5	0.00	0.00	0.00
## 6	0.00	0.06	0.00
## COVID_72_50y_female_NonICU COVID_73_82y_male_NonICU COVID_74_55y_female_ICU			
## 1	0.52	0.46	0.24
## 2	0.00	0.01	0.00
## 3	0.10	0.02	0.12
## 4	0.01	0.02	0.02
## 5	0.00	0.17	0.26
## 6	0.00	0.04	0.00
## COVID_75_55y_male_NonICU COVID_76_73y_female_ICU COVID_77_55y_female_ICU			
## 1	0.23	0.17	0.05
## 2	0.01	0.00	0.00
## 3	0.14	0.09	0.01
## 4	0.00	0.01	0.00
## 5	0.00	0.04	0.00
## 6	0.00	0.00	0.00
## COVID_78_80y_male_NonICU COVID_79_27y_male_NonICU COVID_80_71y_male_ICU			
## 1	0.19	0.08	0.28
## 2	0.00	0.01	0.00
## 3	0.20	0.03	0.05
## 4	0.00	0.00	0.00
## 5	0.00	0.00	0.05
## 6	0.00	0.00	0.00
## COVID_82_67y_male_NonICU COVID_83_85y_female_NonICU			
## 1	0.39	0.47	
## 2	0.01	0.00	
## 3	0.10	0.18	
## 4	0.00	0.05	
## 5	0.00	0.00	
## 6	0.00	0.00	
## COVID_84_75y_female_NonICU COVID_85_62y_male_ICU COVID_86_52y_female_NonICU			

## 1	0.35	0.29	0.60
## 2	0.00	0.00	0.00
## 3	0.03	0.04	0.27
## 4	0.00	0.00	0.02
## 5	0.17	0.00	0.00
## 6	0.00	0.00	0.00
## COVID_87_61y_male_ICU COVID_89_90y_female_NonICU COVID_90_86y_female_NonICU			
## 1	0.65	0.20	0.40
## 2	0.00	0.00	0.00
## 3	0.15	0.07	0.05
## 4	0.00	0.03	0.01
## 5	0.00	0.14	0.31
## 6	0.00	0.00	0.02
## COVID_91_29y_female_NonICU COVID_92_82y_female_ICU COVID_93_81y_female_ICU			
## 1	0.60	0.34	0.37
## 2	0.00	0.00	0.00
## 3	0.03	0.02	0.11
## 4	0.02	0.04	0.00
## 5	0.05	0.58	0.05
## 6	0.00	0.00	0.00
## COVID_94_24y_female_NonICU COVID_95_49y_male_NonICU COVID_96_51y_male_NonICU			
## 1	0.81	0.37	1.61
## 2	0.00	0.01	0.00
## 3	0.17	0.20	0.02
## 4	0.02	0.02	0.00
## 5	0.00	0.15	0.00
## 6	0.06	0.00	0.00
## COVID_97_76y_male_ICU COVID_98_81y_male_NonICU COVID_99_71y_male_ICU			
## 1	0.19	0.78	0.33
## 2	0.00	0.00	0.00
## 3	0.02	0.26	0.02
## 4	0.05	0.00	0.00
## 5	0.12	0.37	0.04
## 6	0.03	0.00	0.00
## COVID_100_74y_female_NonICU COVID_101_58y_male_ICU COVID_102_84y_male_NonICU			
## 1	0.30	0.33	0.12
## 2	0.00	0.00	0.00
## 3	0.09	0.11	0.01
## 4	0.00	0.03	0.01
## 5	0.04	0.05	0.00
## 6	0.00	0.00	0.07
## COVID_103_83y_male_NonICU NONCOVID_01_54y_female_NonICU			
## 1	0.20	0.89	
## 2	0.00	0.00	
## 3	0.03	0.04	
## 4	0.03	0.00	
## 5	0.04	0.00	
## 6	0.00	0.00	
## NONCOVID_02_65y_male_ICU NONCOVID_03_65y_male_ICU NONCOVID_04_90y_male_NonICU			
## 1	0.32	0.44	0.21
## 2	0.00	0.00	0.00
## 3	0.01	0.05	0.05
## 4	0.00	0.02	0.00
## 5	0.04	0.04	0.21

## 6	0.00	0.00	0.00
##	NONCOVID_05_83y_female_NonICU	NONCOVID_06_75y_female_ICU	
## 1	0.31	0.89	
## 2	0.00	0.00	
## 3	0.01	0.14	
## 4	0.01	0.01	
## 5	0.00	0.00	
## 6	0.00	0.06	
##	NONCOVID_07_50y_male_ICU	NONCOVID_08_53y_female_ICU	
## 1	0.45	0.47	
## 2	0.00	0.01	
## 3	0.07	0.04	
## 4	0.02	0.00	
## 5	0.00	0.15	
## 6	0.00	0.00	
##	NONCOVID_09_49y_female_NonICU	NONCOVID_10_67y_male_ICU	
## 1	0.40	0.33	
## 2	0.00	0.00	
## 3	0.04	0.05	
## 4	0.00	0.01	
## 5	0.00	0.23	
## 6	0.00	0.08	
##	NONCOVID_11_58y_female_NonICU	NONCOVID_12_82y_male_ICU	
## 1	0.58	0.12	
## 2	0.00	0.00	
## 3	0.03	0.02	
## 4	0.00	0.00	
## 5	0.00	0.00	
## 6	0.00	0.02	
##	NONCOVID_13_65y_male_ICU	NONCOVID_14_75y_female_ICU	
## 1	0.31	0.16	
## 2	0.00	0.00	
## 3	0.04	0.08	
## 4	0.01	0.00	
## 5	0.32	0.05	
## 6	0.02	0.02	
##	NONCOVID_15_83y_unknown_ICU	NONCOVID_16_40y_female_ICU	
## 1	0.59	0.34	
## 2	0.00	0.00	
## 3	0.03	0.07	
## 4	0.04	0.00	
## 5	0.00	0.13	
## 6	0.19	0.00	
##	NONCOVID_17_84y_female_ICU	NONCOVID_18_88y_male_ICU	
## 1	0.37	0.33	
## 2	0.00	0.00	
## 3	0.07	0.06	
## 4	0.01	0.00	
## 5	0.18	0.00	
## 6	0.00	0.00	
##	NONCOVID_19_66y_female_ICU	NONCOVID_20_62y_female_ICU	
## 1	0.25	0.20	
## 2	0.00	0.00	
## 3	0.11	0.01	

```
## 4          0.00          0.02
## 5          0.04          0.00
## 6          0.03          0.07
##  NONCOVID_21_71y_male_NonICU NONCOVID_22_63y_male_NonICU
## 1          0.40          0.30
## 2          0.00          0.00
## 3          0.04          0.02
## 4          0.02          0.02
## 5          0.00          0.00
## 6          0.00          0.00
##  NONCOVID_23_42y_female_NonICU NONCOVID_24_32y_female_NonICU
## 1          0.70          0.75
## 2          0.00          0.00
## 3          0.02          0.27
## 4          0.01          0.00
## 5          0.00          0.06
## 6          0.00          0.00
##  NONCOVID_25_62y_male_NonICU NONCOVID_26_36y_male_ICU
## 1          2.80          0.22
## 2          0.00          0.00
## 3          0.04          0.28
## 4          0.00          0.00
## 5          0.00          0.00
## 6          0.00          0.00
```

```
#assign first column name as "Gene"
colnames(g1)[1] = 'Gene'
print(head(g1))
```

```
##      Gene COVID_01_39y_male_NonICU COVID_02_63y_male_NonICU
## 1   A1BG          0.49          0.29
## 2   A1CF          0.00          0.00
## 3   A2M           0.21          0.14
## 4  A2ML1          0.04          0.00
## 5 A3GALT2          0.07          0.00
## 6  A4GALT          0.00          0.00
##  COVID_03_33y_male_NonICU COVID_04_49y_male_NonICU COVID_05_49y_male_NonICU
## 1          0.26          0.45          0.17
## 2          0.00          0.01          0.00
## 3          0.03          0.09          0.00
## 4          0.02          0.07          0.05
## 5          0.00          0.00          0.07
## 6          0.00          0.00          0.00
##  COVID_06_.y_male_NonICU COVID_07_38y_female_NonICU COVID_08_78y_male_ICU
## 1          0.21          0.49          0.12
## 2          0.00          0.01          0.00
## 3          0.08          0.23          0.08
## 4          0.04          0.03          0.01
## 5          0.00          0.07          0.00
## 6          0.00          0.00          0.00
##  COVID_09_64y_female_ICU COVID_10_62y_male_ICU COVID_11_52y_female_NonICU
## 1          0.51          0.10          0.38
## 2          0.01          0.00          0.02
## 3          0.88          0.13          0.47
```

## 4	0.02	0.01	0.03
## 5	0.79	0.15	0.08
## 6	0.00	0.00	0.00
## COVID_12_50y_male_ICU	COVID_13_37y_male_NonICU	COVID_14_55y_male_ICU	
## 1	0.45	0.18	0.23
## 2	0.00	0.00	0.00
## 3	0.16	0.07	0.22
## 4	0.00	0.01	0.04
## 5	1.75	0.00	0.93
## 6	0.00	0.00	0.00
## COVID_15_68y_male_ICU	COVID_16_48y_male_NonICU	COVID_17_54y_male_NonICU	
## 1	0.42	0.41	0.63
## 2	0.00	0.01	0.02
## 3	0.07	0.58	0.15
## 4	0.00	0.00	0.02
## 5	0.15	0.19	0.00
## 6	0.03	0.00	0.00
## COVID_18_70y_female_NonICU	COVID_19_51y_male_NonICU	COVID_20_62y_male_ICU	
## 1	0.47	0.33	0.32
## 2	0.00	0.02	0.00
## 3	0.30	0.11	0.07
## 4	0.02	0.02	0.00
## 5	0.06	0.00	0.22
## 6	0.03	0.00	0.00
## COVID_21_66y_male_ICU	COVID_22_43y_male_ICU	COVID_23_76y_male_ICU	
## 1	0.18	0.09	0.18
## 2	0.00	0.00	0.01
## 3	0.00	0.06	0.03
## 4	0.00	0.00	0.00
## 5	0.37	0.06	0.07
## 6	0.03	0.00	0.03
## COVID_24_55y_male_ICU	COVID_25_55y_male_ICU	COVID_26_41y_female_ICU	
## 1	0.22	0.29	0.42
## 2	0.01	0.00	0.00
## 3	0.11	0.09	0.18
## 4	0.02	0.03	0.00
## 5	0.15	0.00	0.87
## 6	0.00	0.00	0.00
## COVID_27_71y_female_ICU	COVID_28_63y_male_ICU	COVID_29_63y_female_ICU	
## 1	0.16	0.18	0.35
## 2	0.01	0.00	0.00
## 3	0.23	0.18	0.03
## 4	0.01	0.05	0.03
## 5	0.18	0.45	0.15
## 6	0.00	0.00	0.03
## COVID_30_54y_male_ICU	COVID_31_50y_male_ICU	COVID_32_72y_male_ICU	
## 1	0.23	0.15	0.34
## 2	0.00	0.00	0.01
## 3	0.11	0.47	0.04
## 4	0.01	0.00	0.00
## 5	0.00	0.00	0.29
## 6	0.00	0.03	0.00
## COVID_33_81y_male_NonICU	COVID_34_64y_female_NonICU		
## 1	0.35	0.36	

## 2	0.00	0.00	
## 3	0.30	0.11	
## 4	0.06	0.00	
## 5	0.26	0.12	
## 6	0.00	0.00	
## COVID_35_58y_female_NonICU	COVID_36_68y_male_NonICU	COVID_37_87y_male_NonICU	
## 1	0.26	0.18	0.20
## 2	0.00	0.01	0.00
## 3	0.51	0.09	0.09
## 4	0.02	0.00	0.07
## 5	0.16	0.08	0.31
## 6	0.00	0.00	0.00
## COVID_38_68y_male_ICU	COVID_39_80y_female_ICU	COVID_40_66y_male_ICU	
## 1	0.29	0.19	0.22
## 2	0.00	0.00	0.00
## 3	0.10	0.27	0.17
## 4	0.02	0.00	0.00
## 5	0.35	0.00	0.08
## 6	0.00	0.07	0.00
## COVID_41_74y_male_ICU	COVID_42_21y_female_ICU	COVID_43_83y_female_ICU	
## 1	0.19	0.24	0.29
## 2	0.00	0.01	0.00
## 3	0.14	0.33	0.00
## 4	0.00	0.01	0.00
## 5	0.19	0.39	0.11
## 6	0.00	0.00	0.00
## COVID_44_46y_male_ICU	COVID_45_62y_female_ICU	COVID_46_62y_male_ICU	
## 1	0.22	0.14	0.53
## 2	0.00	0.00	0.01
## 3	0.14	0.15	0.10
## 4	0.00	0.03	0.00
## 5	0.00	0.19	0.06
## 6	0.04	0.00	0.00
## COVID_47_78y_male_ICU	COVID_48_72y_female_ICU	COVID_49_73y_male_ICU	
## 1	0.08	0.19	0.48
## 2	0.01	0.00	0.00
## 3	0.04	0.06	0.09
## 4	0.03	0.01	0.03
## 5	0.60	0.23	0.00
## 6	0.00	0.06	0.00
## COVID_50_37y_male_ICU	COVID_51_58y_female_NonICU	COVID_52_71y_male_NonICU	
## 1	0.08	0.21	0.25
## 2	0.00	0.00	0.01
## 3	0.01	0.13	0.00
## 4	0.00	0.00	0.03
## 5	0.00	0.00	0.00
## 6	0.72	0.00	0.00
## COVID_53_35y_female_NonICU	COVID_55_62y_female_ICU	COVID_56_33y_female_NonICU	
## 1	0.25	0.09	0.28
## 2	0.00	0.00	0.00
## 3	0.64	0.09	0.16
## 4	0.10	0.01	0.09
## 5	0.00	0.00	0.23
## 6	0.00	0.00	0.00

##	COVID_57_30y_female_NonICU	COVID_58_62y_male_NonICU	COVID_59_55y_male_NonICU
## 1	0.42	0.39	0.33
## 2	0.00	0.00	0.00
## 3	0.27	0.08	0.10
## 4	0.01	0.00	0.00
## 5	0.19	0.00	0.07
## 6	0.05	0.00	0.00
##	COVID_60_49y_male_NonICU	COVID_61_54y_female_NonICU	COVID_62_78y_female_ICU
## 1	0.22	0.25	0.21
## 2	0.00	0.00	0.00
## 3	0.14	0.10	0.04
## 4	0.00	0.03	0.00
## 5	0.00	0.13	0.05
## 6	0.02	0.00	0.00
##	COVID_63_39y_female_ICU	COVID_64_65y_male_ICU	COVID_65_84y_male_NonICU
## 1	0.29	0.38	0.40
## 2	0.00	0.01	0.01
## 3	0.01	0.04	0.07
## 4	0.00	0.02	0.00
## 5	0.14	0.56	0.58
## 6	0.00	0.00	0.00
##	COVID_66_66y_female_NonICU	COVID_67_57y_male_ICU	COVID_68_79y_male_ICU
## 1	0.64	0.37	0.58
## 2	0.00	0.00	0.00
## 3	0.00	0.35	0.15
## 4	0.00	0.00	0.01
## 5	0.00	0.00	0.00
## 6	0.00	0.00	0.05
##	COVID_69_77y_female_NonICU	COVID_70_81y_male_NonICU	COVID_71_37y_male_ICU
## 1	0.52	0.27	0.07
## 2	0.00	0.00	0.01
## 3	0.29	0.07	0.12
## 4	0.02	0.00	0.01
## 5	0.00	0.00	0.00
## 6	0.00	0.06	0.00
##	COVID_72_50y_female_NonICU	COVID_73_82y_male_NonICU	COVID_74_55y_female_ICU
## 1	0.52	0.46	0.24
## 2	0.00	0.01	0.00
## 3	0.10	0.02	0.12
## 4	0.01	0.02	0.02
## 5	0.00	0.17	0.26
## 6	0.00	0.04	0.00
##	COVID_75_55y_male_NonICU	COVID_76_73y_female_ICU	COVID_77_55y_female_ICU
## 1	0.23	0.17	0.05
## 2	0.01	0.00	0.00
## 3	0.14	0.09	0.01
## 4	0.00	0.01	0.00
## 5	0.00	0.04	0.00
## 6	0.00	0.00	0.00
##	COVID_78_80y_male_NonICU	COVID_79_27y_male_NonICU	COVID_80_71y_male_ICU
## 1	0.19	0.08	0.28
## 2	0.00	0.01	0.00
## 3	0.20	0.03	0.05
## 4	0.00	0.00	0.00

## 5	0.00	0.00	0.05
## 6	0.00	0.00	0.00
## COVID_82_67y_male_NonICU	COVID_83_85y_female_NonICU		
## 1	0.39	0.47	
## 2	0.01	0.00	
## 3	0.10	0.18	
## 4	0.00	0.05	
## 5	0.00	0.00	
## 6	0.00	0.00	
## COVID_84_75y_female_NonICU	COVID_85_62y_male_ICU	COVID_86_52y_female_NonICU	
## 1	0.35	0.29	0.60
## 2	0.00	0.00	0.00
## 3	0.03	0.04	0.27
## 4	0.00	0.00	0.02
## 5	0.17	0.00	0.00
## 6	0.00	0.00	0.00
## COVID_87_61y_male_ICU	COVID_89_90y_female_NonICU	COVID_90_86y_female_NonICU	
## 1	0.65	0.20	0.40
## 2	0.00	0.00	0.00
## 3	0.15	0.07	0.05
## 4	0.00	0.03	0.01
## 5	0.00	0.14	0.31
## 6	0.00	0.00	0.02
## COVID_91_29y_female_NonICU	COVID_92_82y_female_ICU	COVID_93_81y_female_ICU	
## 1	0.60	0.34	0.37
## 2	0.00	0.00	0.00
## 3	0.03	0.02	0.11
## 4	0.02	0.04	0.00
## 5	0.05	0.58	0.05
## 6	0.00	0.00	0.00
## COVID_94_24y_female_NonICU	COVID_95_49y_male_NonICU	COVID_96_51y_male_NonICU	
## 1	0.81	0.37	1.61
## 2	0.00	0.01	0.00
## 3	0.17	0.20	0.02
## 4	0.02	0.02	0.00
## 5	0.00	0.15	0.00
## 6	0.06	0.00	0.00
## COVID_97_76y_male_ICU	COVID_98_81y_male_NonICU	COVID_99_71y_male_ICU	
## 1	0.19	0.78	0.33
## 2	0.00	0.00	0.00
## 3	0.02	0.26	0.02
## 4	0.05	0.00	0.00
## 5	0.12	0.37	0.04
## 6	0.03	0.00	0.00
## COVID_100_74y_female_NonICU	COVID_101_58y_male_ICU	COVID_102_84y_male_NonICU	
## 1	0.30	0.33	0.12
## 2	0.00	0.00	0.00
## 3	0.09	0.11	0.01
## 4	0.00	0.03	0.01
## 5	0.04	0.05	0.00
## 6	0.00	0.00	0.07
## COVID_103_83y_male_NonICU	NONCOVID_01_54y_female_NonICU		
## 1	0.20	0.89	
## 2	0.00	0.00	

## 3	0.03	0.04	
## 4	0.03	0.00	
## 5	0.04	0.00	
## 6	0.00	0.00	
##	NONCOVID_02_65y_male_ICU	NONCOVID_03_65y_male_ICU	NONCOVID_04_90y_male_NonICU
## 1	0.32	0.44	0.21
## 2	0.00	0.00	0.00
## 3	0.01	0.05	0.05
## 4	0.00	0.02	0.00
## 5	0.04	0.04	0.21
## 6	0.00	0.00	0.00
##	NONCOVID_05_83y_female_NonICU	NONCOVID_06_75y_female_ICU	
## 1	0.31	0.89	
## 2	0.00	0.00	
## 3	0.01	0.14	
## 4	0.01	0.01	
## 5	0.00	0.00	
## 6	0.00	0.06	
##	NONCOVID_07_50y_male_ICU	NONCOVID_08_53y_female_ICU	
## 1	0.45	0.47	
## 2	0.00	0.01	
## 3	0.07	0.04	
## 4	0.02	0.00	
## 5	0.00	0.15	
## 6	0.00	0.00	
##	NONCOVID_09_49y_female_NonICU	NONCOVID_10_67y_male_ICU	
## 1	0.40	0.33	
## 2	0.00	0.00	
## 3	0.04	0.05	
## 4	0.00	0.01	
## 5	0.00	0.23	
## 6	0.00	0.08	
##	NONCOVID_11_58y_female_NonICU	NONCOVID_12_82y_male_ICU	
## 1	0.58	0.12	
## 2	0.00	0.00	
## 3	0.03	0.02	
## 4	0.00	0.00	
## 5	0.00	0.00	
## 6	0.00	0.02	
##	NONCOVID_13_65y_male_ICU	NONCOVID_14_75y_female_ICU	
## 1	0.31	0.16	
## 2	0.00	0.00	
## 3	0.04	0.08	
## 4	0.01	0.00	
## 5	0.32	0.05	
## 6	0.02	0.02	
##	NONCOVID_15_83y_unknown_ICU	NONCOVID_16_40y_female_ICU	
## 1	0.59	0.34	
## 2	0.00	0.00	
## 3	0.03	0.07	
## 4	0.04	0.00	
## 5	0.00	0.13	
## 6	0.19	0.00	
##	NONCOVID_17_84y_female_ICU	NONCOVID_18_88y_male_ICU	

```

## 1          0.37          0.33
## 2          0.00          0.00
## 3          0.07          0.06
## 4          0.01          0.00
## 5          0.18          0.00
## 6          0.00          0.00
##  NONCOVID_19_66y_female_ICU NONCOVID_20_62y_female_ICU
## 1          0.25          0.20
## 2          0.00          0.00
## 3          0.11          0.01
## 4          0.00          0.02
## 5          0.04          0.00
## 6          0.03          0.07
##  NONCOVID_21_71y_male_NonICU NONCOVID_22_63y_male_NonICU
## 1          0.40          0.30
## 2          0.00          0.00
## 3          0.04          0.02
## 4          0.02          0.02
## 5          0.00          0.00
## 6          0.00          0.00
##  NONCOVID_23_42y_female_NonICU NONCOVID_24_32y_female_NonICU
## 1          0.70          0.75
## 2          0.00          0.00
## 3          0.02          0.27
## 4          0.01          0.00
## 5          0.00          0.06
## 6          0.00          0.00
##  NONCOVID_25_62y_male_NonICU NONCOVID_26_36y_male_ICU
## 1          2.80          0.22
## 2          0.00          0.00
## 3          0.04          0.28
## 4          0.00          0.00
## 5          0.00          0.00
## 6          0.00          0.00

```

```
print(head(sm1))
```

```

##           participant_id geo_accession          status
## 1 COVID_01_39y_male_NonICU   GSM4753021 Public on Aug 29 2020
## 2 COVID_02_63y_male_NonICU   GSM4753022 Public on Aug 29 2020
## 3 COVID_03_33y_male_NonICU   GSM4753023 Public on Aug 29 2020
## 4 COVID_04_49y_male_NonICU   GSM4753024 Public on Aug 29 2020
## 5 COVID_05_49y_male_NonICU   GSM4753025 Public on Aug 29 2020
## 6 COVID_06_:y_male_NonICU    GSM4753026 Public on Aug 29 2020
##  X.Sample_submission_date last_update_date type channel_count
## 1          Aug 28 2020      Aug 29 2020  SRA             1
## 2          Aug 28 2020      Aug 29 2020  SRA             1
## 3          Aug 28 2020      Aug 29 2020  SRA             1
## 4          Aug 28 2020      Aug 29 2020  SRA             1
## 5          Aug 28 2020      Aug 29 2020  SRA             1
## 6          Aug 28 2020      Aug 29 2020  SRA             1
##           source_name_ch1 organism_ch1      disease_status age  sex
## 1 Leukocytes from whole blood Homo sapiens disease state: COVID-19 39  male
## 2 Leukocytes from whole blood Homo sapiens disease state: COVID-19 63  male

```

```

## 3 Leukocytes from whole blood Homo sapiens disease state: COVID-19 33 male
## 4 Leukocytes from whole blood Homo sapiens disease state: COVID-19 49 male
## 5 Leukocytes from whole blood Homo sapiens disease state: COVID-19 49 male
## 6 Leukocytes from whole blood Homo sapiens disease state: COVID-19 : male
## icu_status apacheii charlson_score mechanical_ventilation
## 1 no 15 0 yes
## 2 no unknown 2 no
## 3 no unknown 2 no
## 4 no unknown 1 no
## 5 no 19 1 yes
## 6 no unknown 1 no
## ventilator.free_days hospital.free_days_post_45_day_followup ferritin.ng.ml.
## 1 0 0 946
## 2 28 39 1060
## 3 28 18 1335
## 4 28 39 583
## 5 23 27 800
## 6 28 36 563
## crp.mg.l. ddimer.mg.l_feu. procalcitonin.ng.ml.. lactate.mmol.l. fibrinogen
## 1 73.1 1.3 36 0.9 513
## 2 unknown 1.03 0.37 unknown unknown
## 3 53.2 1.48 0.07 unknown 513
## 4 251.1 1.32 0.98 0.87 949
## 5 355.8 0.69 4.92 1.48 929
## 6 129.1 unknown 0.67 0.86 769
## sofa
## 1 8
## 2 unknown
## 3 unknown
## 4 unknown
## 5 7
## 6 unknown

```

```
gene_sel = 'A1CF' #2
```

```
#extract gene data from that row
```

```
grow = g1[g1$Gene == gene_sel,-1] #select rows where Gene is A1CF
```

```
print(grow)
```

```

## COVID_01_39y_male_NonICU COVID_02_63y_male_NonICU COVID_03_33y_male_NonICU
## 2 0 0 0
## COVID_04_49y_male_NonICU COVID_05_49y_male_NonICU COVID_06_.y_male_NonICU
## 2 0.01 0 0
## COVID_07_38y_female_NonICU COVID_08_78y_male_ICU COVID_09_64y_female_ICU
## 2 0.01 0 0.01
## COVID_10_62y_male_ICU COVID_11_52y_female_NonICU COVID_12_50y_male_ICU
## 2 0 0.02 0
## COVID_13_37y_male_NonICU COVID_14_55y_male_ICU COVID_15_68y_male_ICU
## 2 0 0 0
## COVID_16_48y_male_NonICU COVID_17_54y_male_NonICU COVID_18_70y_female_NonICU
## 2 0.01 0.02 0
## COVID_19_51y_male_NonICU COVID_20_62y_male_ICU COVID_21_66y_male_ICU
## 2 0.02 0 0
## COVID_22_43y_male_ICU COVID_23_76y_male_ICU COVID_24_55y_male_ICU

```

```

## 2          0          0.01          0.01
## COVID_25_55y_male_ICU COVID_26_41y_female_ICU COVID_27_71y_female_ICU
## 2          0          0          0.01
## COVID_28_63y_male_ICU COVID_29_63y_female_ICU COVID_30_54y_male_ICU
## 2          0          0          0
## COVID_31_50y_male_ICU COVID_32_72y_male_ICU COVID_33_81y_male_NonICU
## 2          0          0.01          0
## COVID_34_64y_female_NonICU COVID_35_58y_female_NonICU
## 2          0          0
## COVID_36_68y_male_NonICU COVID_37_87y_male_NonICU COVID_38_68y_male_ICU
## 2          0.01          0          0
## COVID_39_80y_female_ICU COVID_40_66y_male_ICU COVID_41_74y_male_ICU
## 2          0          0          0
## COVID_42_21y_female_ICU COVID_43_83y_female_ICU COVID_44_46y_male_ICU
## 2          0.01          0          0
## COVID_45_62y_female_ICU COVID_46_62y_male_ICU COVID_47_78y_male_ICU
## 2          0          0.01          0.01
## COVID_48_72y_female_ICU COVID_49_73y_male_ICU COVID_50_37y_male_ICU
## 2          0          0          0
## COVID_51_58y_female_NonICU COVID_52_71y_male_NonICU
## 2          0          0.01
## COVID_53_35y_female_NonICU COVID_55_62y_female_ICU COVID_56_33y_female_NonICU
## 2          0          0          0
## COVID_57_30y_female_NonICU COVID_58_62y_male_NonICU COVID_59_55y_male_NonICU
## 2          0          0          0
## COVID_60_49y_male_NonICU COVID_61_54y_female_NonICU COVID_62_78y_female_ICU
## 2          0          0          0
## COVID_63_39y_female_ICU COVID_64_65y_male_ICU COVID_65_84y_male_NonICU
## 2          0          0.01          0.01
## COVID_66_66y_female_NonICU COVID_67_57y_male_ICU COVID_68_79y_male_ICU
## 2          0          0          0
## COVID_69_77y_female_NonICU COVID_70_81y_male_NonICU COVID_71_37y_male_ICU
## 2          0          0          0.01
## COVID_72_50y_female_NonICU COVID_73_82y_male_NonICU COVID_74_55y_female_ICU
## 2          0          0.01          0
## COVID_75_55y_male_NonICU COVID_76_73y_female_ICU COVID_77_55y_female_ICU
## 2          0.01          0          0
## COVID_78_80y_male_NonICU COVID_79_27y_male_NonICU COVID_80_71y_male_ICU
## 2          0          0.01          0
## COVID_82_67y_male_NonICU COVID_83_85y_female_NonICU
## 2          0.01          0
## COVID_84_75y_female_NonICU COVID_85_62y_male_ICU COVID_86_52y_female_NonICU
## 2          0          0          0
## COVID_87_61y_male_ICU COVID_89_90y_female_NonICU COVID_90_86y_female_NonICU
## 2          0          0          0
## COVID_91_29y_female_NonICU COVID_92_82y_female_ICU COVID_93_81y_female_ICU
## 2          0          0          0
## COVID_94_24y_female_NonICU COVID_95_49y_male_NonICU COVID_96_51y_male_NonICU
## 2          0          0.01          0
## COVID_97_76y_male_ICU COVID_98_81y_male_NonICU COVID_99_71y_male_ICU
## 2          0          0          0
## COVID_100_74y_female_NonICU COVID_101_58y_male_ICU COVID_102_84y_male_NonICU
## 2          0          0          0
## COVID_103_83y_male_NonICU NONCOVID_01_54y_female_NonICU

```

```
## 2          0          0
##  NONCOVID_02_65y_male_ICU NONCOVID_03_65y_male_ICU NONCOVID_04_90y_male_NonICU
## 2          0          0          0
##  NONCOVID_05_83y_female_NonICU NONCOVID_06_75y_female_ICU
## 2          0          0
##  NONCOVID_07_50y_male_ICU NONCOVID_08_53y_female_ICU
## 2          0          0.01
##  NONCOVID_09_49y_female_NonICU NONCOVID_10_67y_male_ICU
## 2          0          0
##  NONCOVID_11_58y_female_NonICU NONCOVID_12_82y_male_ICU
## 2          0          0
##  NONCOVID_13_65y_male_ICU NONCOVID_14_75y_female_ICU
## 2          0          0
##  NONCOVID_15_83y_unknown_ICU NONCOVID_16_40y_female_ICU
## 2          0          0
##  NONCOVID_17_84y_female_ICU NONCOVID_18_88y_male_ICU
## 2          0          0
##  NONCOVID_19_66y_female_ICU NONCOVID_20_62y_female_ICU
## 2          0          0
##  NONCOVID_21_71y_male_NonICU NONCOVID_22_63y_male_NonICU
## 2          0          0
##  NONCOVID_23_42y_female_NonICU NONCOVID_24_32y_female_NonICU
## 2          0          0
##  NONCOVID_25_62y_male_NonICU NONCOVID_26_36y_male_ICU
## 2          0          0
```

```
#transpose
growtp = data.frame(participant_id = colnames(g1)[-1],
expression = as.numeric(grow))
#convert to data frame with pid and expression(of gene A1CF)

print(head(growtp))
```

```
##      participant_id expression
## 1 COVID_01_39y_male_NonICU    0.00
## 2 COVID_02_63y_male_NonICU    0.00
## 3 COVID_03_33y_male_NonICU    0.00
## 4 COVID_04_49y_male_NonICU    0.01
## 5 COVID_05_49y_male_NonICU    0.00
## 6 COVID_06_.y_male_NonICU     0.00
```

```
#add metadata by merge w sm1
mg = merge(growtp,sm1,by='participant_id') #merge by participant_id
print(head(mg))
```

```
##      participant_id expression geo_accession      status
## 1 COVID_01_39y_male_NonICU    0.00 GSM4753021 Public on Aug 29 2020
## 2 COVID_02_63y_male_NonICU    0.00 GSM4753022 Public on Aug 29 2020
## 3 COVID_03_33y_male_NonICU    0.00 GSM4753023 Public on Aug 29 2020
## 4 COVID_04_49y_male_NonICU    0.01 GSM4753024 Public on Aug 29 2020
## 5 COVID_05_49y_male_NonICU    0.00 GSM4753025 Public on Aug 29 2020
## 6 COVID_07_38y_female_NonICU  0.01 GSM4753027 Public on Aug 29 2020
## X.Sample_submission_date last_update_date type channel_count
```



```

## 1      Aug 28 2020      Aug 29 2020  SRA      1
## 2      Aug 28 2020      Aug 29 2020  SRA      1
## 3      Aug 28 2020      Aug 29 2020  SRA      1
## 4      Aug 28 2020      Aug 29 2020  SRA      1
## 5      Aug 28 2020      Aug 29 2020  SRA      1
## 6      Aug 28 2020      Aug 29 2020  SRA      1
##      source_name_ch1 organism_ch1      disease_status age      sex
## 1 Leukocytes from whole blood Homo sapiens disease state: COVID-19 39      male
## 2 Leukocytes from whole blood Homo sapiens disease state: COVID-19 63      male
## 3 Leukocytes from whole blood Homo sapiens disease state: COVID-19 33      male
## 4 Leukocytes from whole blood Homo sapiens disease state: COVID-19 49      male
## 5 Leukocytes from whole blood Homo sapiens disease state: COVID-19 49      male
## 6 Leukocytes from whole blood Homo sapiens disease state: COVID-19 38      female
##      icu_status apacheii charlson_score mechanical_ventilation
## 1      no      15      0      yes
## 2      no      unknown      2      no
## 3      no      unknown      2      no
## 4      no      unknown      1      no
## 5      no      19      1      yes
## 6      no      unknown      7      no
##      ventilator.free_days hospital.free_days_post_45_day_followup ferritin.ng.ml.
## 1      0      0      946
## 2      28      39      1060
## 3      28      18      1335
## 4      28      39      583
## 5      23      27      800
## 6      28      42      366
##      crp.mg.l. ddimer.mg.l_feu. procalcitonin.ng.ml.. lactate.mmol.l. fibrinogen
## 1      73.1      1.3      36      0.9      513
## 2      unknown      1.03      0.37      unknown      unknown
## 3      53.2      1.48      0.07      unknown      513
## 4      251.1      1.32      0.98      0.87      949
## 5      355.8      0.69      4.92      1.48      929
## 6      unknown      0.87      0.06      1.17      478
##      sofa
## 1      8
## 2      unknown
## 3      unknown
## 4      unknown
## 5      7
## 6      unknown

```

```

#Gene: A1CF
#ContinuousCovariate: Age (numeric)
#CategoricalCovariate: sex (male/female)
#CategoricalCovariate: mechanical_ventilation (yes/no)

#1 histogram of gene expression (for A1CF)

library(ggplot2)
summary(mg$expression)

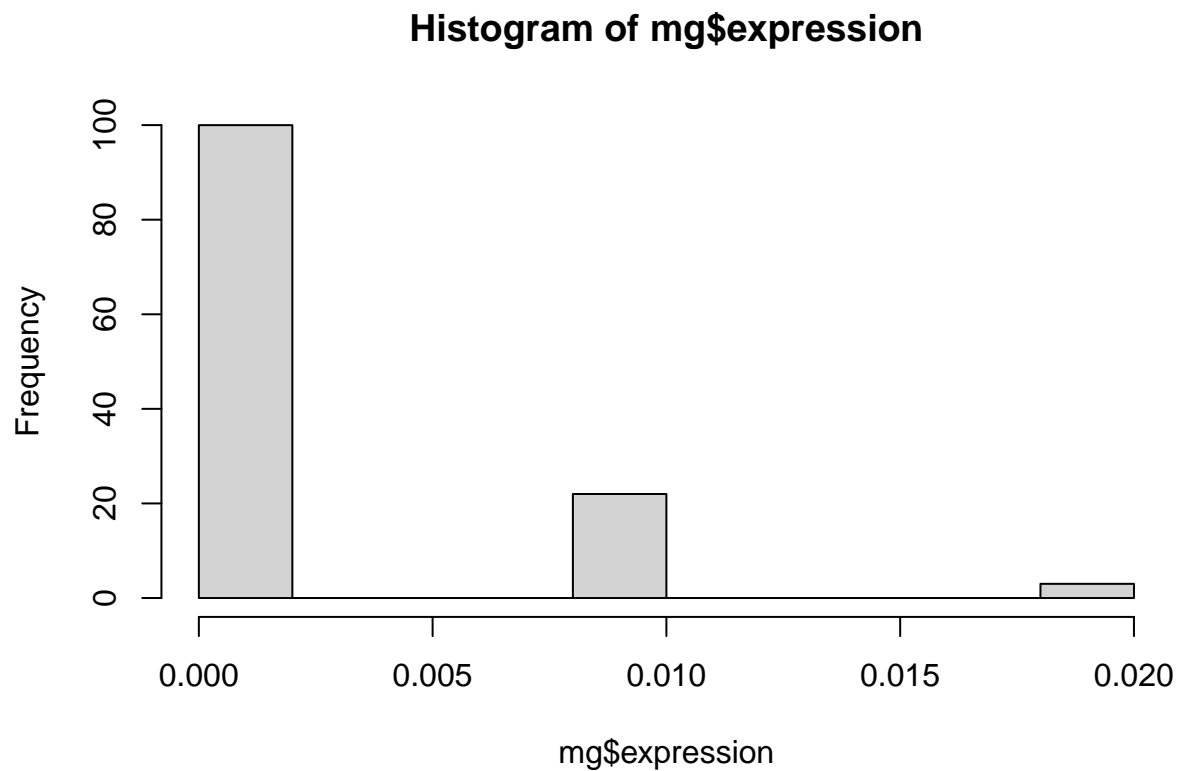
```

```

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.00000 0.00000 0.00224 0.00000 0.02000

```

```
hist(mg$expression) #duh
```



```
#2 Scatterplot of Gene Expression and Age
```

```
ggplot(mg, aes(x = age, y = expression)) + #merged data, expression vals vs age vals
```

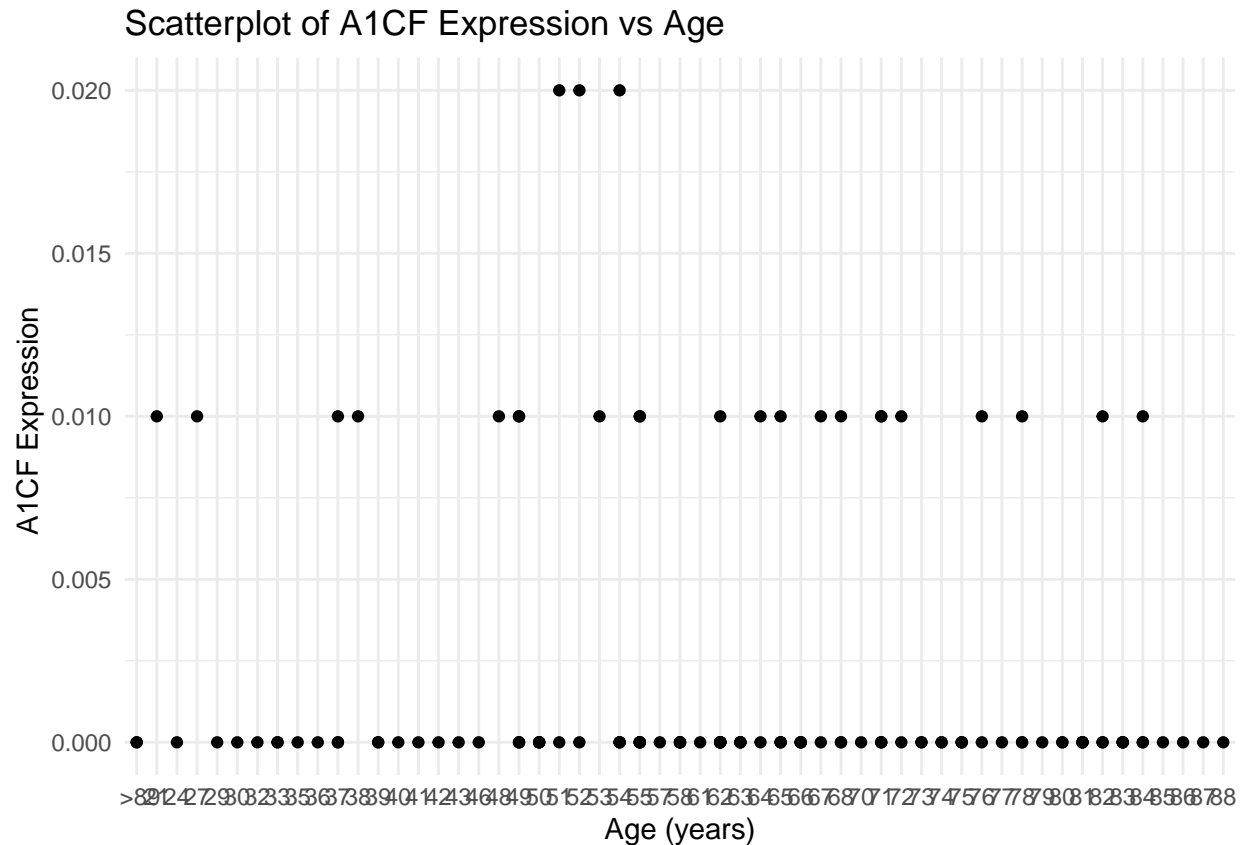
```
  geom_point() +
```

```
  labs(title = paste("Scatterplot of", gene_sel, "Expression vs Age"),
```

```
        x = "Age (years)",
```

```
        y = paste(gene_sel, "Expression ")) + #likely log2-ed for value range, compresses a bit for view
```

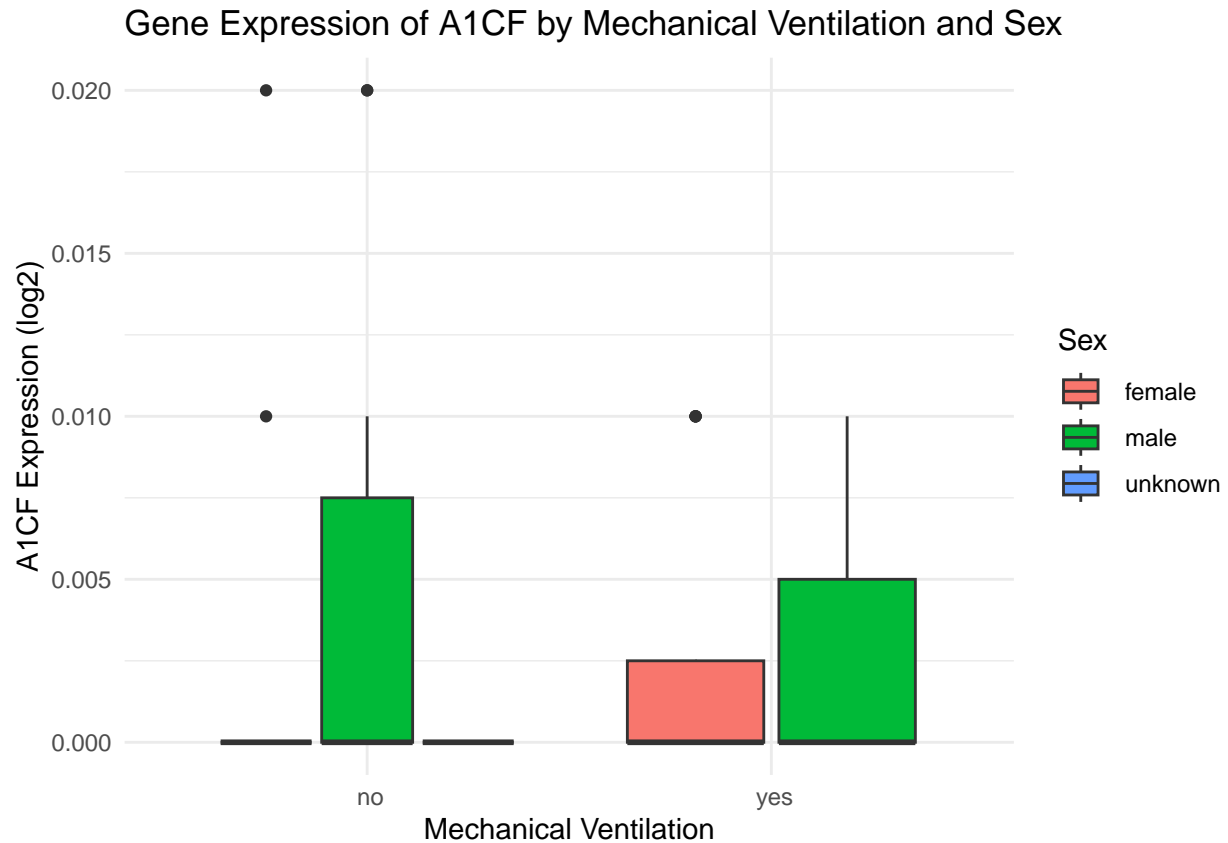
```
  theme_minimal()
```



*#explanation: Gene A1CF and Covariate Age: doesnt seem to be a strong relationship visually between a
#and the A1CF expression values, although the data is a little bit sparse.
#A1CF expressions are somewhat quantized at 0 and .01 (log2). definitely more values at 0.
more analysis of other genes, covariates needed. no linear regression here,
#but fairly visually obvious a lack of correlation*

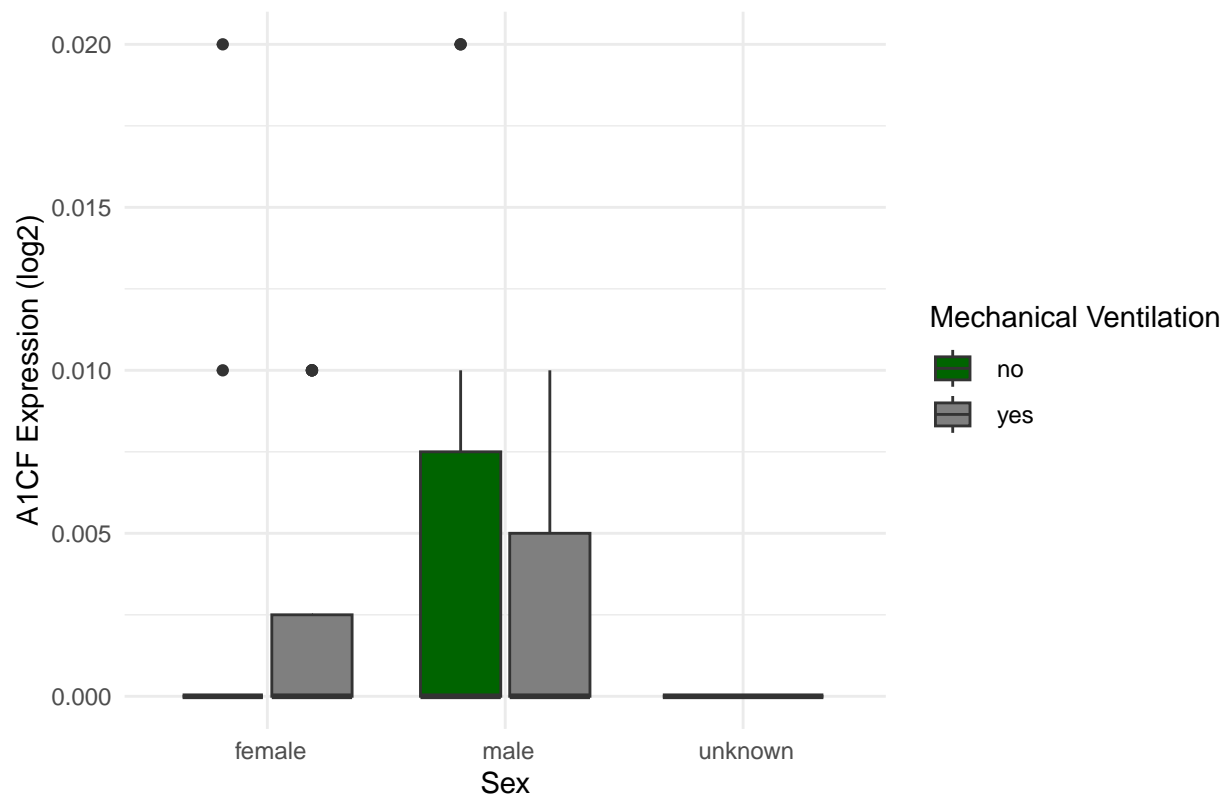
```
# boxplot params opt: outlier.shape = 21, outlier.size = 2, alpha = 0.8
#3: Boxplot of gene expression
#separated by both categorical covariates (sex, mechanical_ventilation(1st))
ggplot(mg, aes(x = mechanical_ventilation, y = expression, fill = sex)) +
  geom_boxplot() +

labs(
  title = "Gene Expression of A1CF by Mechanical Ventilation and Sex",
  x = "Mechanical Ventilation", #primary x separation
  y = "A1CF Expression (log2)",
  fill = "Sex" #secondary separation
) +
theme_minimal()
```



```
#3: optional reverse of categorical nesting levels (dif interp potentially )
ggplot(mg, aes(x = sex, y = expression, fill = mechanical_ventilation)) +
  geom_boxplot() +
  scale_fill_manual(values = c("darkgreen", "grey50")) +
  labs(
    title = "Gene Expression of A1CF by Sex and Mechanical Ventilation",
    x = "Sex", #swapped w mechanical_ventilation from above
    y = "A1CF Expression (log2)",
    fill = "Mechanical Ventilation"
  ) +
  theme_minimal()
```

Gene Expression of A1CF by Sex and Mechanical Ventilation



#explanation:

```
print(mg$age)
```

```
## [1] "39" "63" "33" "49" "49" "38" "78" "64" "62" "74"
## [11] "58" "84" "83" "52" "50" "37" "55" "68" "48" "54"
## [21] "70" "51" "62" "66" "43" "76" "55" "55" "41" "71"
## [31] "63" "63" "54" "50" "72" "81" "64" "58" "68" "87"
## [41] "68" "80" "66" "74" "21" "83" "46" "62" "62" "78"
## [51] "72" "73" "37" "58" "71" "35" "62" "33" "30" "62"
## [61] "55" "49" "54" "78" "39" "65" "84" "66" "57" "79"
## [71] "77" "81" "37" "50" "82" "55" "55" "73" "55" "80"
## [81] "27" "71" "67" "85" "75" "62" "52" "61" " >89" "86"
## [91] "29" "82" "81" "24" "49" "51" "76" "81" "71" "54"
## [101] "65" "65" " >89" "83" "75" "50" "53" "49" "67" "58"
## [111] "82" "65" "75" "83" "40" "84" "88" "66" "62" "71"
## [121] "63" "42" "32" "62" "36"
```

End of part 1. knit boio