

DSC180A: TEAM 2 final_result.pdf

Members: Jiyeon Song , Vibha Sastry , Oakkar Aung , Rihana Mohamed

Table of Contents:

- analysis_notebooks/
 - [02_kmeans.ipynb](#)
 - [03_Ica.Rmd](#)
 - [04_morbidity_composition.ipynb](#)
 - [05_logistic_regression.R](#)
- eda_notebooks/
 - [02_network_visualizations.ipynb](#)
 - [03_circular_barplots.ipynb](#)
 - [04_boxplot_sub_composition.ipynb](#)

analysis_notebooks/

02_kmeans.ipynb

Cluster 0

```
disease_prevelance[disease_prevelance['cluster'] == 0]
```

	group_1	group_2	group_3	group_4	group_5	cluster
valvular_disease	0.015496	0.034096	0.070937	0.139131	0.187264	0
pulmonary_circulation	0.018595	0.034584	0.051233	0.065666	0.082786	0
peripheral_vascular	0.012397	0.021189	0.052658	0.107019	0.097313	0
paralysis	0.014463	0.027521	0.027754	0.026345	0.029453	0
other_neurological	0.089876	0.122991	0.110096	0.093584	0.103284	0
chronic_pulmonary	0.078512	0.106917	0.170803	0.246805	0.222090	0
diabetes_uncomplicated	0.013430	0.063322	0.178518	0.247264	0.191841	0
diabetes_complicated	0.007231	0.033609	0.058779	0.064618	0.038806	0
hypothyroidism	0.009298	0.040429	0.073118	0.110623	0.161791	0
renal_failure	0.014463	0.042864	0.085444	0.150141	0.203980	0
liver_disease	0.050620	0.116415	0.150344	0.059571	0.027861	0
peptic_ulcer	0.004132	0.004384	0.008553	0.008651	0.011144	0
aids	0.000000	0.013151	0.011488	0.001049	0.000398	0
lymphoma	0.001033	0.010960	0.010565	0.015466	0.013333	0
metastatic_cancer	0.003099	0.025329	0.066577	0.060948	0.037214	0
solid_tumor	0.001033	0.006576	0.025910	0.036569	0.030448	0
rheumatoid_arthritis	0.008264	0.016318	0.025071	0.037093	0.039801	0
coagulopathy	0.052686	0.099854	0.112947	0.103480	0.091741	0
obesity	0.013430	0.049927	0.073201	0.048103	0.007761	0
weight_loss	0.012397	0.027277	0.038236	0.039059	0.042985	0
blood_loss_anemia	0.004132	0.009985	0.017441	0.019988	0.024876	0
deficiency_anemias	0.022727	0.028008	0.020963	0.022413	0.041592	0
alcohol_abuse	0.170455	0.167560	0.138269	0.041156	0.007363	0

Cluster 1

```
: disease_prevelance[disease_prevelance['cluster'] == 1]
```

	group_1	group_2	group_3	group_4	group_5	cluster
congestive_heart_failure	0.027893	0.064540	0.156046	0.297923	0.431841	1
cardiac_arrhythmias	0.091942	0.116659	0.213902	0.438364	0.566965	1
hypertension	0.035124	0.187043	0.476103	0.655023	0.683781	1
fluid_electrolyte	0.176653	0.263517	0.271088	0.283832	0.336517	1

Cluster 2

```
: disease_prevelance[disease_prevelance['cluster'] == 2]
```

	group_1	group_2	group_3	group_4	group_5	cluster
admission_type_encoded	0.972107	0.891622	0.811085	0.806737	0.928358	2

03_lca.Rmd

```
## Chi-square test for congestive_heart_failure
##
## Pearson's Chi-squared test
##
## data: contingency_table
## X-squared = 12823, df = 6, p-value < 2.2e-16
##
```

This is what the expected chi-square test will return. The p-value is less than 0.05 so significant. However in the provided LCA markdown the following will indicate it was run correctly:

All p-values are significant (less than 0.05)

There is sufficient evidence to conclude all associations between each latent class and comorbidity are significant

04_morbidity_composition.ipynb

Condition	Subgroup 1	Subgroup 2	Subgroup 3	Subgroup 4	Subgroup 5	Subgroup 6	sub1_num	sub2_num	sub3_num	sub4_num	sub5_num	sub6_num
Number of patients	2216	8540	3583	3428	9019	9604						
congestive heart failure	73.42	3.43	11.14	51.46	7.81	36.53	congestive_heart_failure →	75.05	0.74	12.41	44.38	7.60 → 49.07
cardiac arrhythmias	62.41	9.34	14.37	48.57	23.12	57.40	cardiac_arrhythmias	63.17	4.66	15.42	35.54	23.68 → 65.81
valvular disease	36.91	1.35	2.71	14.67	8.56	15.75	valvular_disease	38.05	0.76	2.88	11.66	6.55 → 19.75
pulmonary circulation	66.29	2.35	3.85	0.70	0.62	2.11	pulmonary_circulation →	67.10	1.35	3.71	0.98	1.07 → 0.92
peripheral vascular	14.17	1.53	2.40	20.13	9.87	7.00	peripheral_vascular	14.77	0.86	2.60	20.14	7.83 → 11.45
hypertension	64.67	9.16	38.93	89.47	87.70	47.14	hypertension	65.92	4.12	41.45 → 81.77 → 93.53 → 66.71		
paralysis	0.68	3.20	1.95	2.16	1.76	3.86	paralysis	0.92	2.98	2.67	1.35	3.01 → 3.30
other neurological	6.41	12.06	17.44	10.47	2.96	13.61	other_neurological	7.51	11.70	18.61	8.84	7.03 → 11.40
chronic pulmonary	93.86	9.18	17.22	15.61	15.67	18.78	chronic_pulmonary	92.79	5.76	17.54	14.61	15.48 → 17.83
diabetes uncomplicated	28.88	1.02	18.64	21.94	36.82	16.51	diabetes_uncomplicated	28.62	0.78	18.43	0.00 → 34.19	24.78
diabetes complicated	7.67	2.14	2.46	35.41	3.03	1.02	diabetes_complicated	6.64	1.10	2.13 → 98.16 → 0.00		0.29
hypothyroidism	14.21	4.16	5.78	15.11	9.17	11.06	hypothyroidism	14.42	1.25	6.59	12.65	7.70 → 15.70
renal failure	23.06	0.29	7.98	88.30	3.28	0.99	renal_failure	26.61	0.45	10.59	56.29	5.85 → 21.88
liver disease	6.45	2.39	67.01	7.56	1.19	3.29	liver_disease	6.16	4.82 → 60.47	6.14	1.54	3.33
peptic ulcer	0.54	0.27	2.29	1.11	0.58	0.93	peptic_ulcer	0.61	0.18	2.02	0.68	0.60 → 1.15
aids	0.09	0.71	3.80	0.47	0.00	0.00	aids	0.17	0.92	3.71	0.37	0.05 → 0.00
lymphoma	1.26	0.88	1.67	1.28	0.29	2.50	lymphoma	1.18	0.12	1.97	0.74	0.57 → 1.91
metastatic cancer	2.21	7.45	4.47	3.56	3.66	7.47	metastatic_cancer	3.41	3.12	5.92	2.58	5.34 → 4.27
solid tumor	1.90	1.30	5.89	2.36	1.20	4.83	solid_tumor	1.79	0.00	5.87	1.84	2.07 → 3.76
rheumatoid arthritis	5.32	1.41	2.65	4.29	2.34	4.27	rheumatoid_arthritis	5.46	0.67	2.88	3.01	2.44 → 4.64
coagulopathy	14.35	2.87	41.81	13.80	1.21	11.96	coagulopathy	14.29	2.63	40.28	8.10	2.54 → 12.74
obesity	15.88	1.38	5.30	7.15	10.93	0.02	obesity	14.37	1.67	5.04	11.54	9.39 → 0.35
weight loss	3.47	1.99	11.16	4.38	0.00	5.76	weight_loss	3.98	0.96	12.07	3.07	0.18 → 5.75
fluid electrolyte	32.99	19.29	58.39	45.45	7.15	34.96	fluid_electrolyte	35.34	18.07	59.67	37.20	13.64 → 39.12
blood loss anemia	1.76	1.11	3.77	2.60	0.49	2.56	blood_loss_anemia	1.79	0.37	3.94	2.21	0.76 → 2.74
deficiency anemias	5.91	1.99	3.96	4.14	1.03	1.88	deficiency_anemias	5.46	2.21	4.18	2.89	0.93 → 3.64
alcohol abuse	4.96	12.48	47.75	1.28	1.67	1.51	alcohol_abuse	4.72 → 19.42 → 44.93			1.35	2.75 → 0.48
drug abuse	2.62	8.43	18.22	1.55	0.62	0.36	drug_abuse	2.62	15.25	14.48	1.72	1.13 → 0.44
psychoses	1.85	2.11	2.85	1.78	0.63	1.31	psychoses	1.57	2.19	3.58	2.21	0.95 → 1.10
depression	12.50	11.17	20.12	10.04	6.43	4.65	depression	12.10	14.58	19.28	10.99	7.10 → 6.88

05_logistic_regression.R

```
[1] "Logistic Regression for pulmonary_circulation"
Call:
glm(formula = formula, family = binomial(link = "logit"), data = all_data)

Deviance Residuals:
    Min      1Q      Median      3Q      Max 
-1.4912 -0.1647 -0.1466 -0.1361  3.0615 

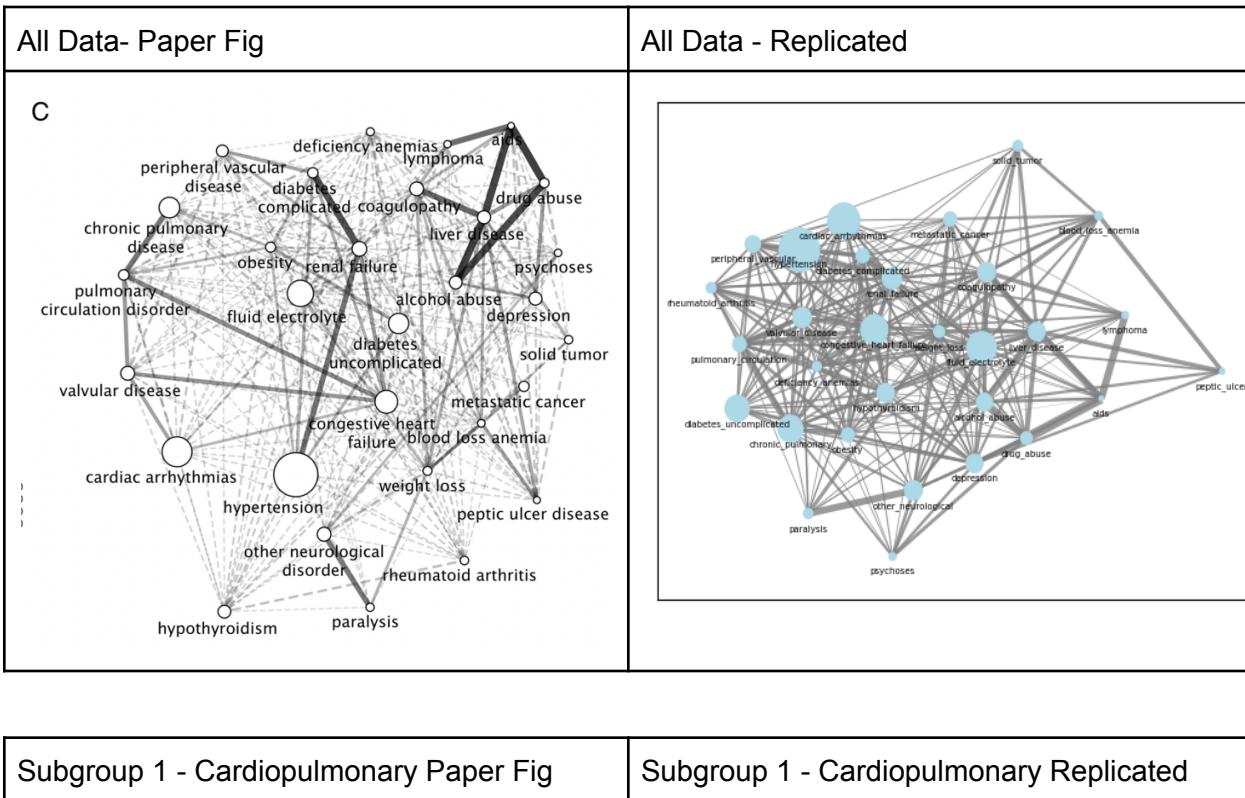
Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) -4.52829   0.09812 -46.153 <2e-16 ***
factor(subgroup)2 -0.14898   0.14711 -1.013   0.311    
factor(subgroup)3  0.23514   0.15807  1.488   0.137    
factor(subgroup)4  5.24116   0.10773  48.651 <2e-16 ***
factor(subgroup)6 -0.08497   0.26971 -0.315   0.753    
factor(subgroup)7  1.27235   0.12996  9.790 <2e-16 ***  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

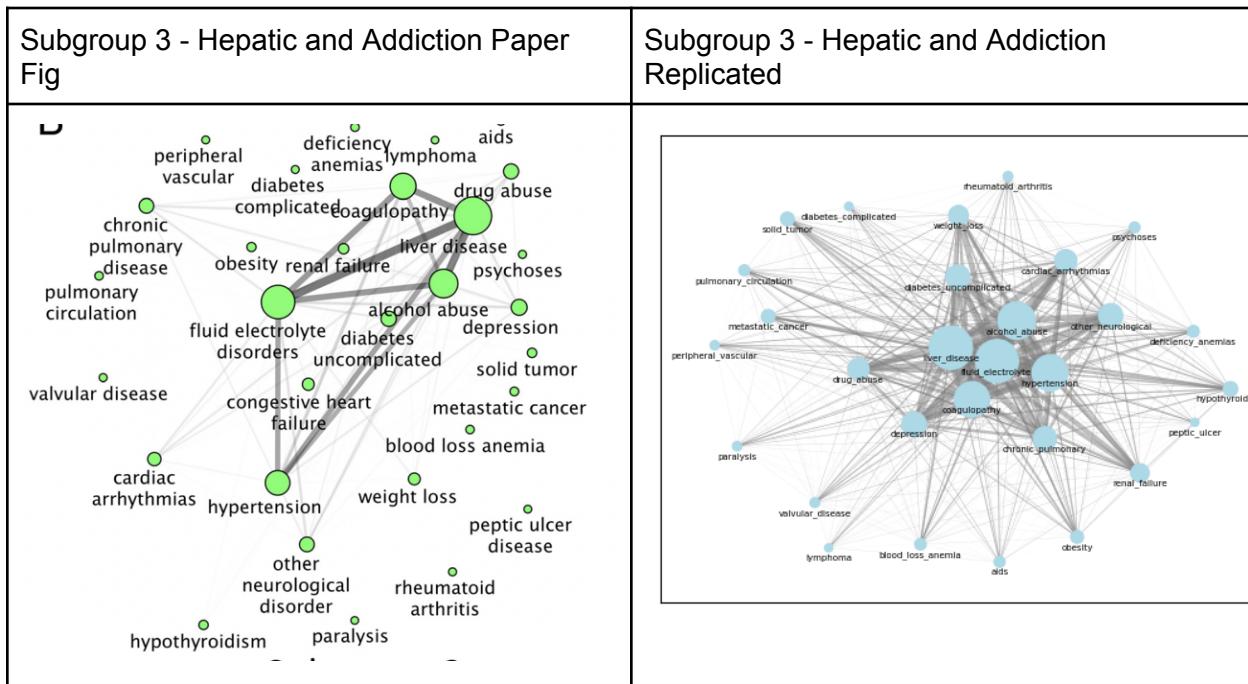
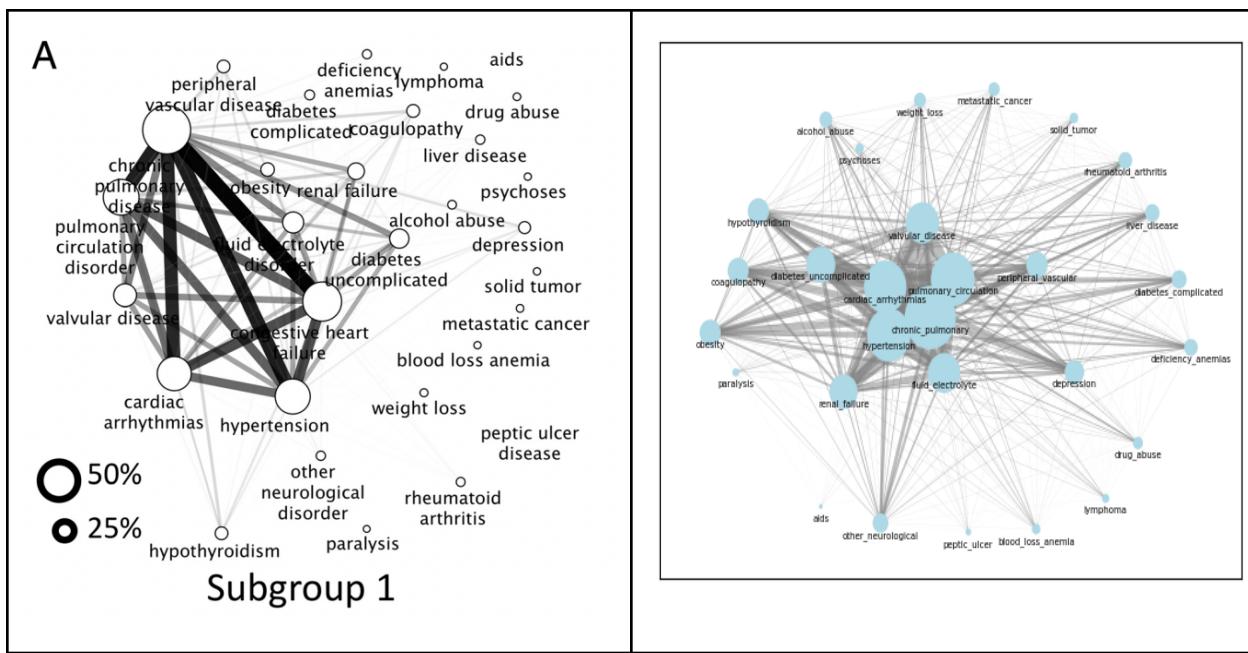
(Dispersion parameter for binomial family taken to be 1)
```

RESULT: pulmonary_circulation is significant in group 1(uncomplicated diabetes group), group 4(cardiopulmonary group), group 7(hepatic & addiction group) of my dataset.

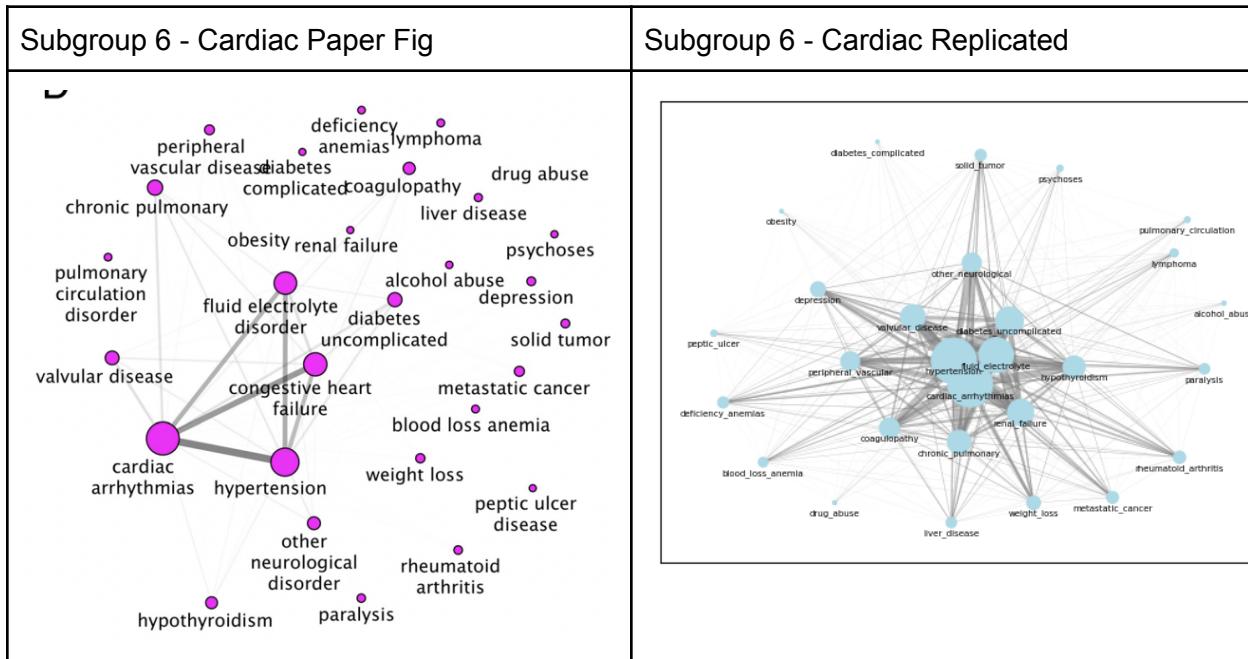
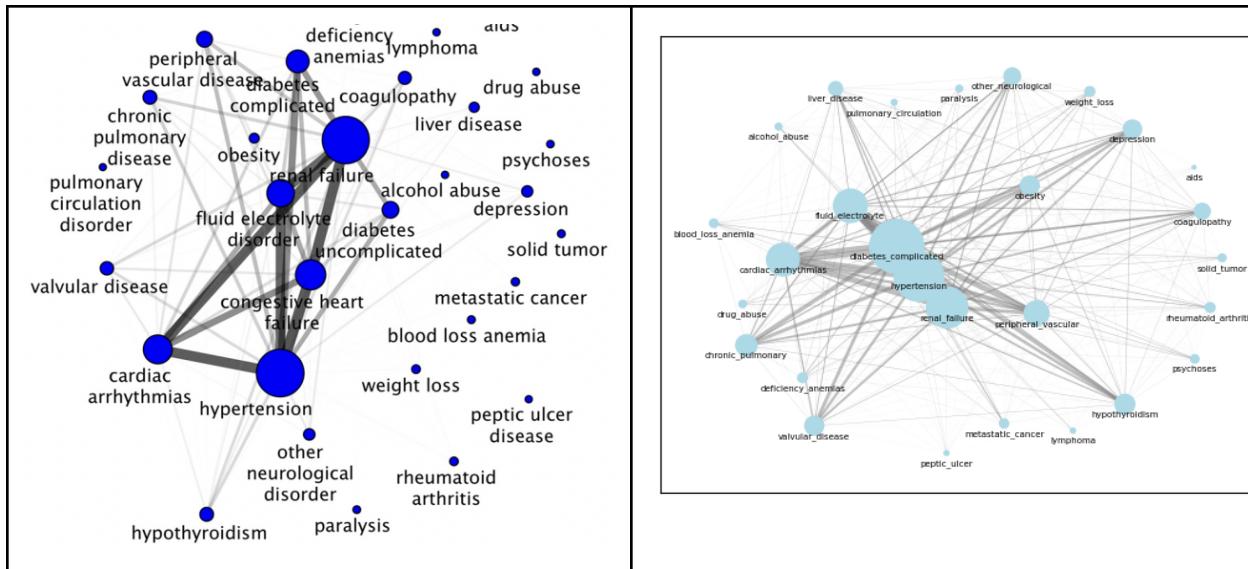
eda_notebooks/

02_network_visualizations.ipynb



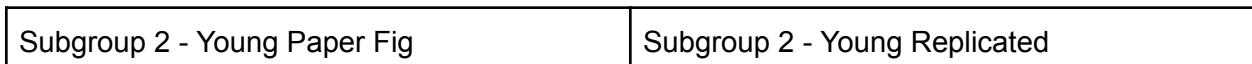
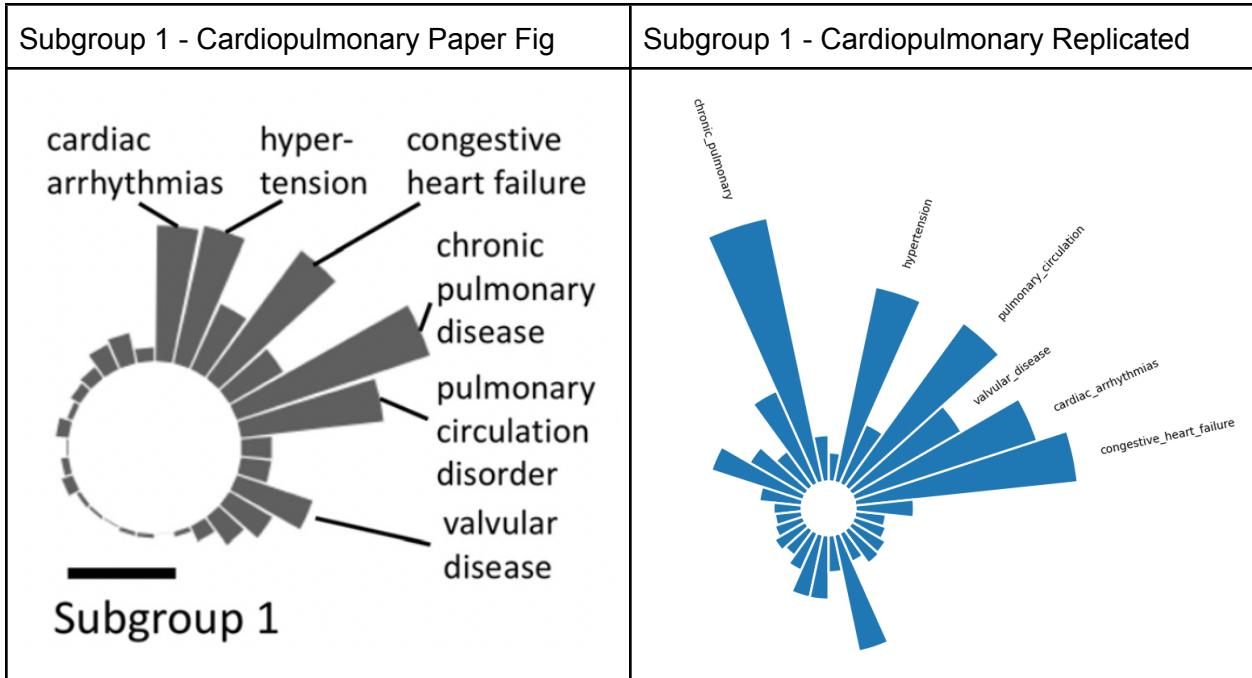
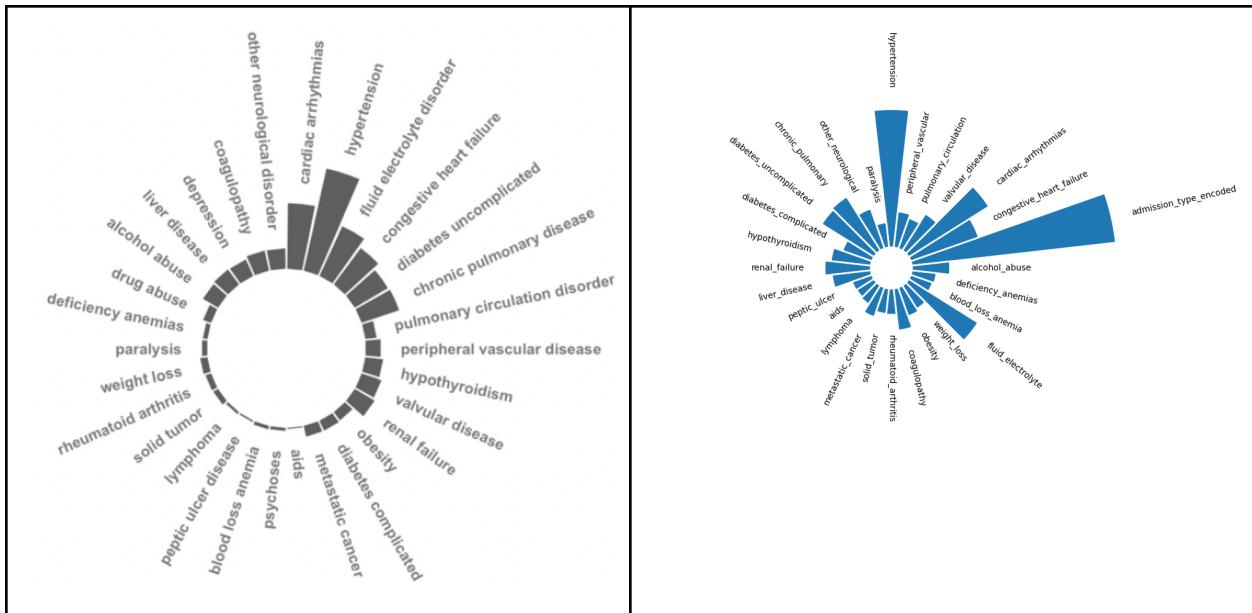


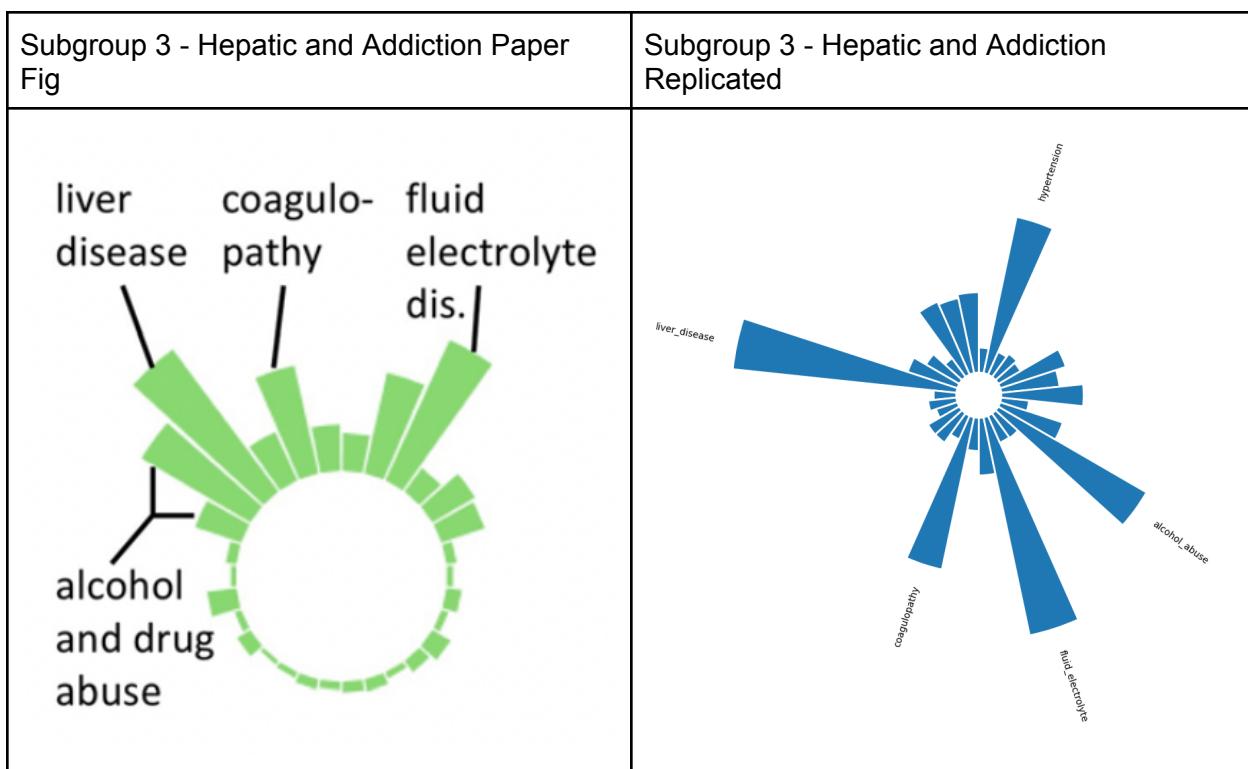
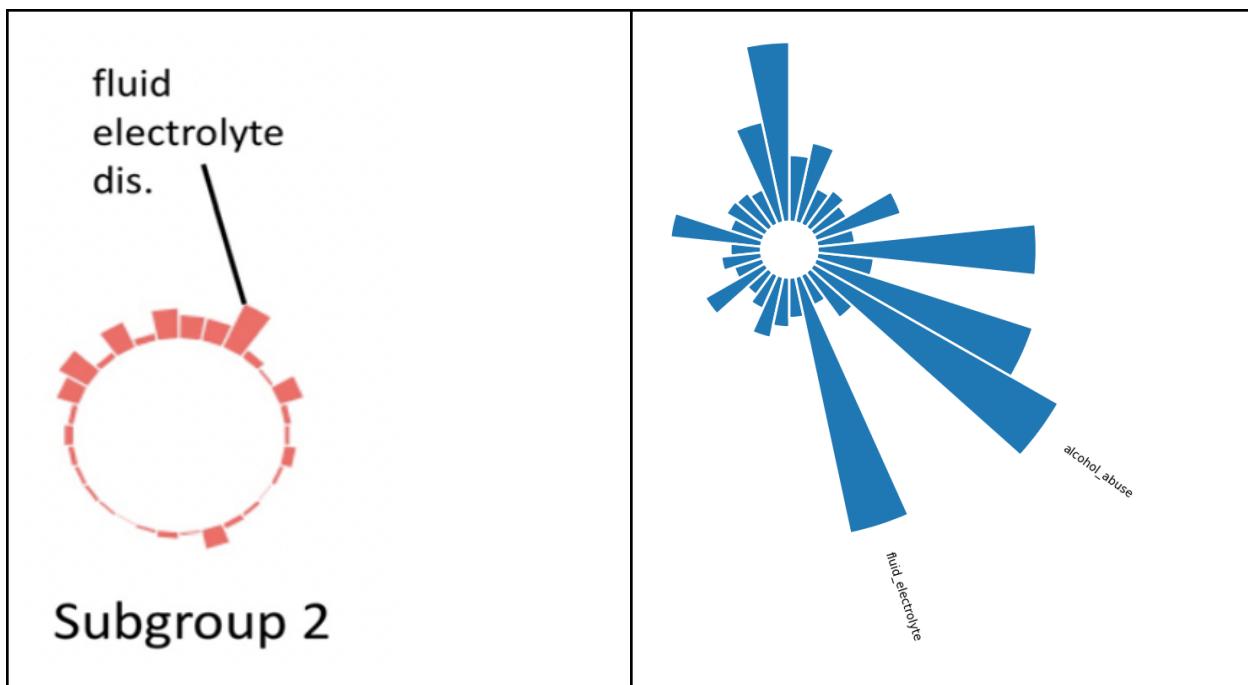
Subgroup 4 - Complicated Diabetes Paper Fig	Subgroup 4 - Complicated Diabetes Replicated
--	---



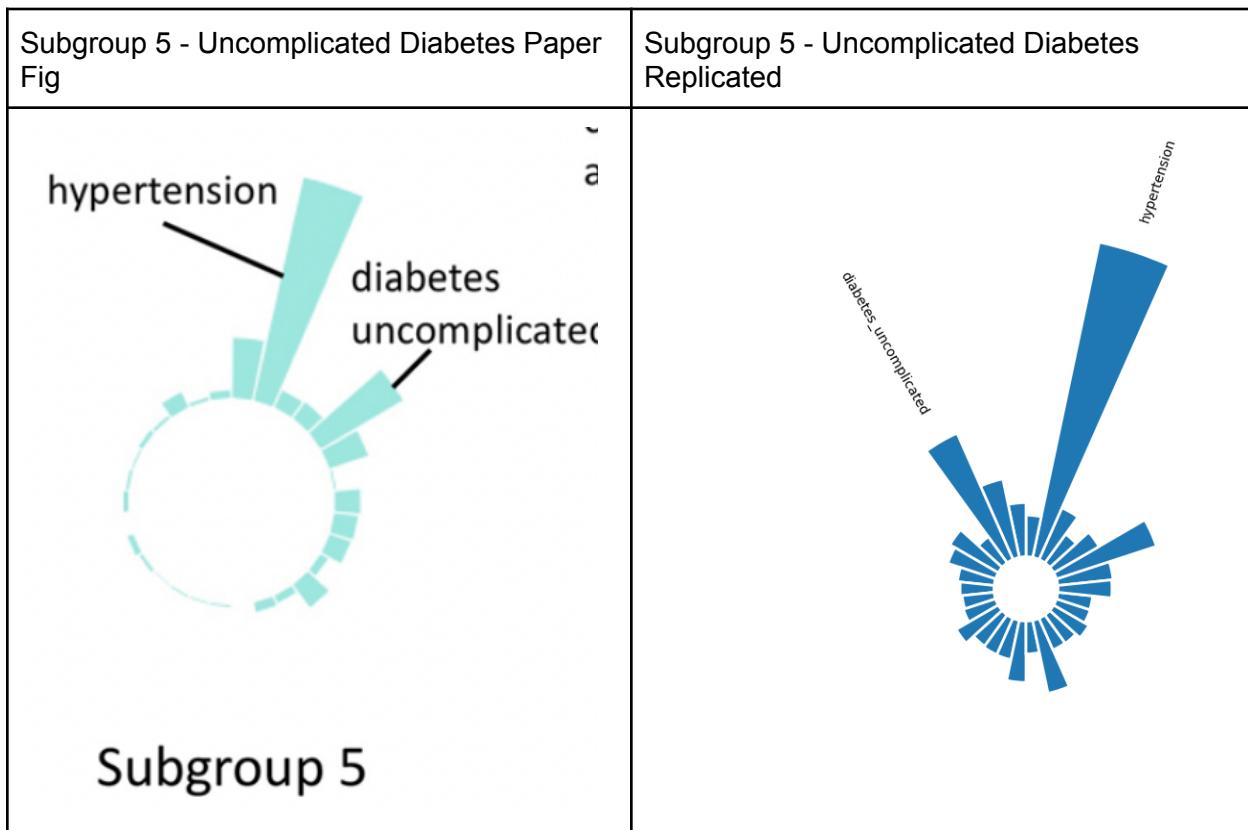
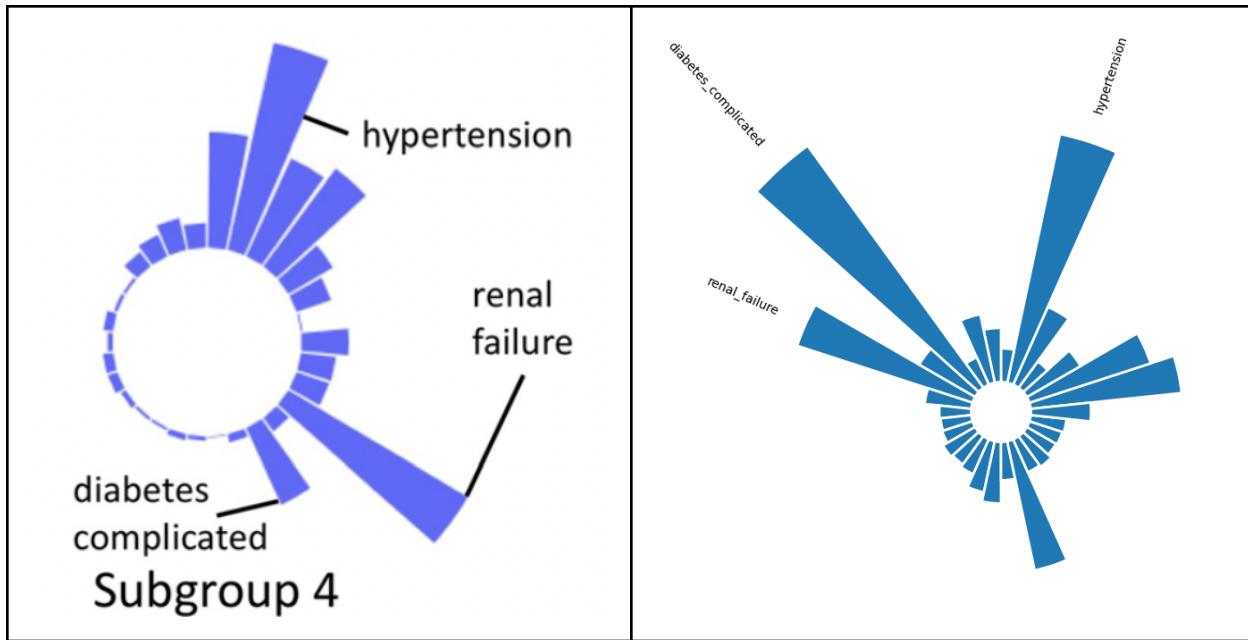
03_circular_barplots.ipynb

All Data- Paper Fig	All Data - Replicated
---------------------	-----------------------

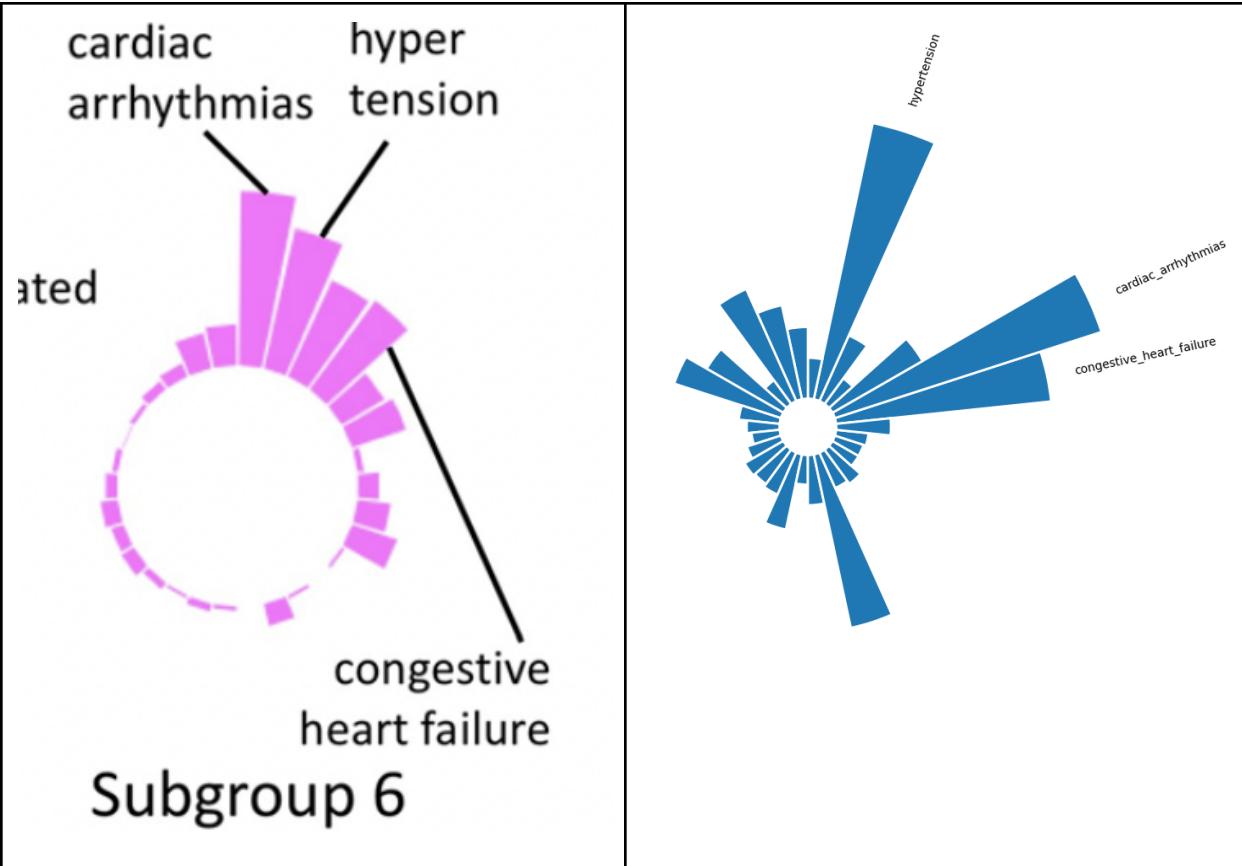




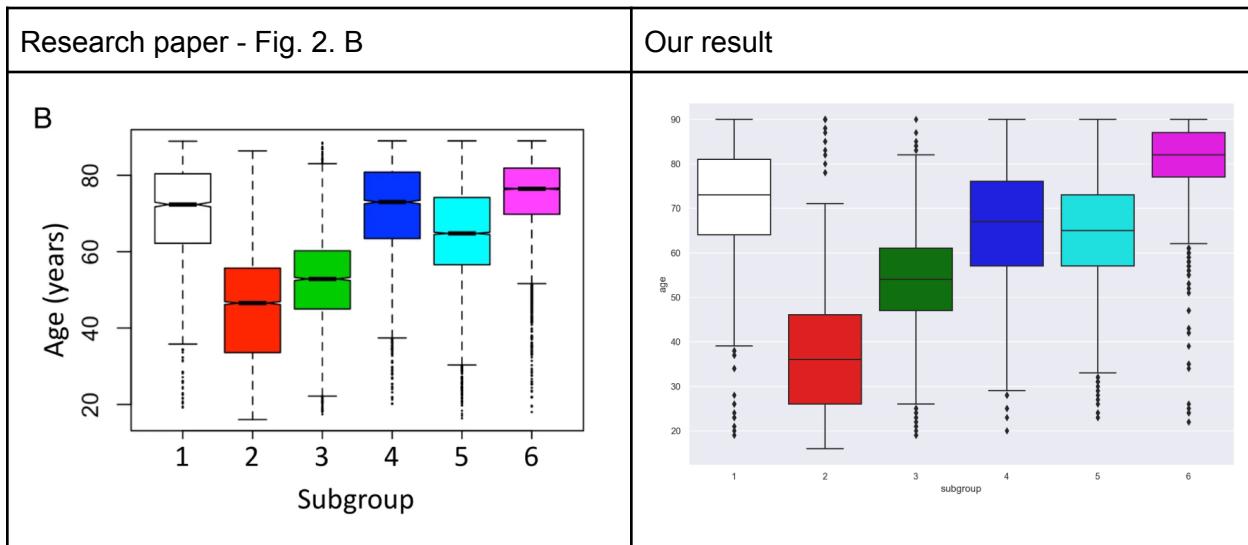
<p>Subgroup 4 - Complicated Diabetes Paper Fig</p>	<p>Subgroup 4 - Complicated Diabetes Replicated</p>
--	---



Subgroup 6 - Cardiac Paper Fig	Subgroup 6 - Cardiac Replicated
--------------------------------	---------------------------------



04_boxplot_sub_age.ipynb



Mean & median age analysis for each subgroup

Mean ages for each subgroup	Median ages for each subgroup
<pre>print(sub1['age'].mean()) print(sub2['age'].mean()) print(sub3['age'].mean()) print(sub4['age'].mean()) print(sub5['age'].mean()) print(sub6['age'].mean()) ✓ 0.0s</pre> <p>71.84097859327217 36.61772513783949 54.32468206592266 65.88275015346838 64.8019943019943 81.09339332748024</p>	<pre>print(sub1['age'].median()) print(sub2['age'].median()) print(sub3['age'].median()) print(sub4['age'].median()) print(sub5['age'].median()) print(sub6['age'].median()) ✓ 0.0s</pre> <p>73.0 36.0 54.0 67.0 65.0 82.0</p>

