The focus of my research is on how health and lifestyle influence sleep in an individual's life. The "Sleep Health and Lifestyle" dataset sourced from Kaggle [1] consists of 12 variables about individual's sleeping habits and related activities. These variables include gender, age group, lifestyle factors (e.g., physical activity level, stress levels, overall activeness, occupation), and sleep-related metrics (such as sleep quality and duration). The dataset also includes health-related data, such as heart rate, BMI, and blood pressure. This dataset has been categorized for better analysis.

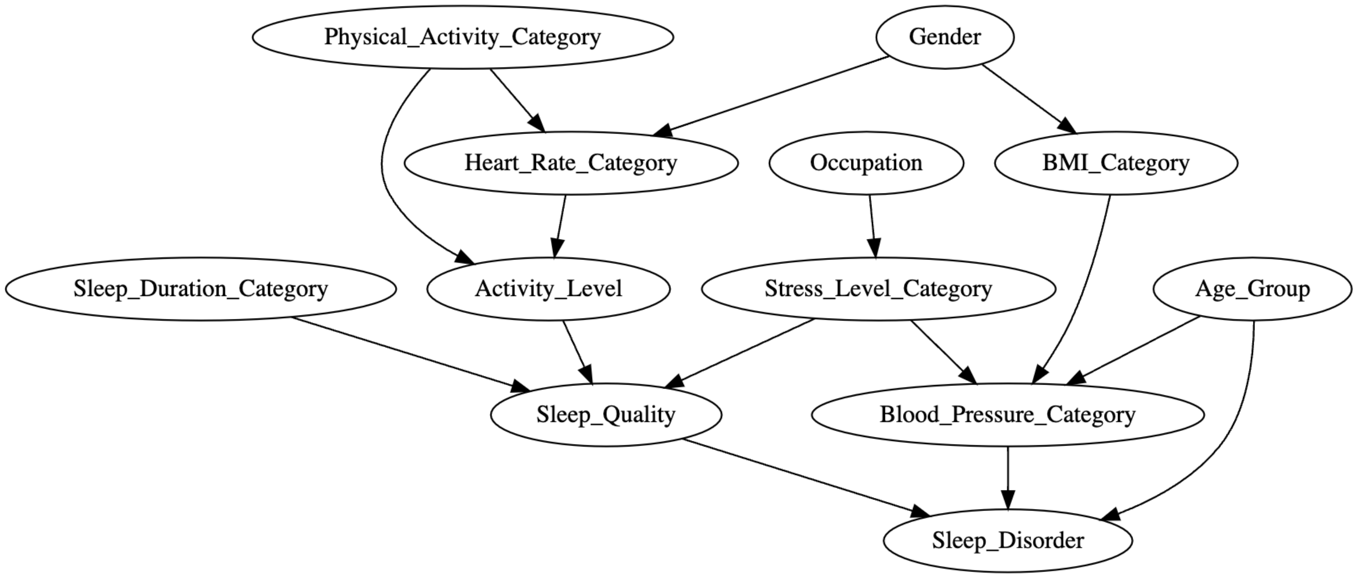
	A	B	C	D	E	F	G	H	I	J	K	L
1	Gender	Age_Group	Occupation	Sleep_Duration_Category	Sleep_Quality	Physical_Activity_Category	Stress_Level_Category	BMI_Category	Blood_Pressure_Category	Heart_Rate_Category	Activity_Level	Sleep_Disorder
2	Male	27-34	Software Engineer	Normal (6-8 hours)	Fair	Low	Moderate Stress	Overweight	Hypertension Stage 1	Normal (71-80 bpm)	Sedentary (<5000 steps)	None
3	Male	27-34	Doctor	Normal (6-8 hours)	Fair	Medium	High Stress	Normal	Hypertension Stage 1	Normal (71-80 bpm)	Very Active (>=10000 steps)	None
4	Male	27-34	Doctor	Normal (6-8 hours)	Fair	Medium	High Stress	Normal	Hypertension Stage 1	Normal (71-80 bpm)	Very Active (>=10000 steps)	None
5	Male	27-34	Sales Representative	Short (<6 hours)	Poor	Low	High Stress	Obese	Hypertension Stage 2	High (81-86 bpm)	Sedentary (<5000 steps)	Sleep Apnea
6	Male	27-34	Sales Representative	Short (<6 hours)	Poor	Low	High Stress	Obese	Hypertension Stage 2	High (81-86 bpm)	Sedentary (<5000 steps)	Sleep Apnea
7	Male	27-34	Software Engineer	Short (<6 hours)	Poor	Low	High Stress	Obese	Hypertension Stage 2	High (81-86 bpm)	Sedentary (<5000 steps)	Insomnia
8	Male	27-34	Teacher	Normal (6-8 hours)	Fair	Low	High Stress	Obese	Hypertension Stage 2	High (81-86 bpm)	Sedentary (<5000 steps)	Insomnia
9	Male	27-34	Doctor	Normal (6-8 hours)	Good	Medium	Moderate Stress	Normal	Hypertension Stage 1	Low (65-70 bpm)	Moderately Active (7500-9999 steps)	None
10	Male	27-34	Doctor	Normal (6-8 hours)	Good	Medium	Moderate Stress	Normal	Hypertension Stage 1	Low (65-70 bpm)	Moderately Active (7500-9999 steps)	None

This data is suitable for structure learning because with a wide range of variables, allowing us to identify patterns and relation between the variables. By the structure algorithms, I hope to discover, what factors affect our sleep health. This analysis by the algorithms can provide insights into the complex interactions between habits and overall health, potentially encouraging healthier sleep and lifestyle choices. One insight may be how occupation, stress combined can influence the quality of sleep.

The knowledge-based DAG depicts how every variable corelates various health and lifestyle factors to reveal their collective influence on sleep. To create this graph, I have mostly used my knowledge, and refer few literatures to obtain knowlegde.

Considering how an individual Occupation can lead to varying levels of stress, which might impact sleep quality and potentially contribute to sleep disorders. [2] The duration and quality of sleep are closely linked, not getting enough sleep can lead to problems with how well we sleep, which is recognized in sleep studies.

Physical activity is another important factor. People who are more active tend to have a healthier heart rate, which can lead to better sleep and a lower risk of sleep disorders. [3] This relationship is well-known in health and fitness research. [4] Weight, as represented by BMI, can influence blood pressure, which is connected to sleep disorders in the diagram. This reflects medical understanding that weight management is crucial for maintaining healthy blood pressure levels, which in turn can influence sleep quality. [5] Aging is associated with several changes in sleep patterns, and as people grow older, they often encounter a range of sleep disorders like Insomnia, sleep apnea, sleep behaviour disorder. Based on some factors listed above I created the DAG graph.



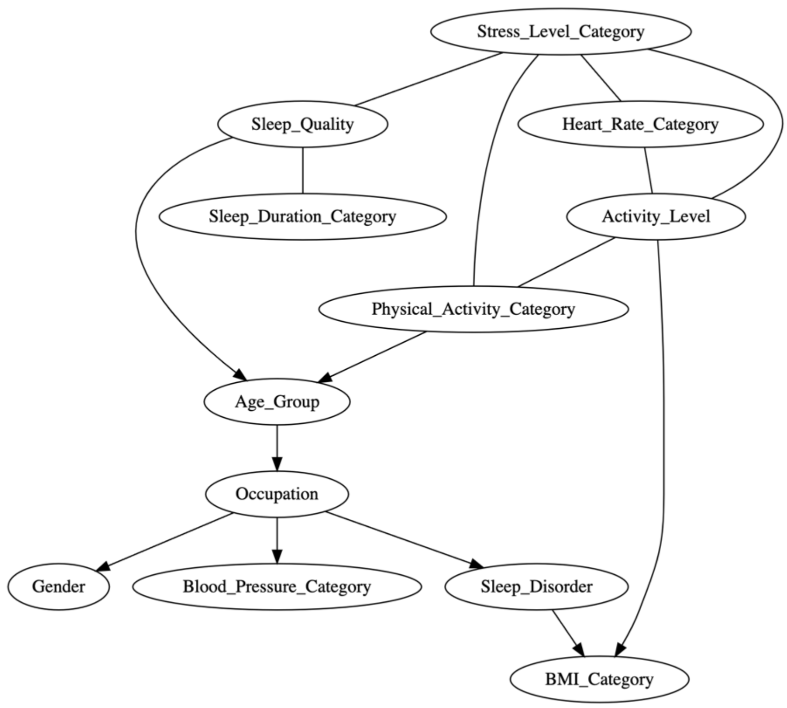
Algorithm	CPDAG scores			Log-Likelihood (LL) score	BIC score	#free parameters	Structure learning elapsed time
	BSF	SHD	F1				
HC	-0.082	24.0	0.133	-3291.824	-4496.936	282	0
TABU	-0.082	24.0	0.133	-3291.824	-4496.936	282	0
SaiyanH	-0.161	28.0	0.118	-3206.532	-4445.832	290	0
MAHC	-0.082	24.0	0.138	-3589.812	-4547.064	224	0
GES	-0.082	24.0	0.133	-3291.824	-4496.936	282	0

When comparing the CPDAG scores with those listed in Table 3.1, I've noticed that my scores are significantly lower. The values I provided in my DAGtrue table are backed by real-world knowledge, and the values given in the table are based on the observed patterns specific to their use case. This also indicates that the algorithms fit to my data is less than ideal with measured against the DAG table. The negative BSF scores, which compared with the positive scores in the manual, imply the difference between the algo assumptions and the actual dependencies in my data.

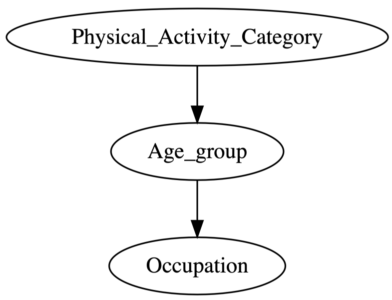
The SHD score, mostly at 24 with a exception of 28 for SaiyanH, are higher than some of the lower SHD scores found in the manual. This might imply a greater difference between the learned structure and the true DAG, which could be a result of your data's unique complexities or difference in how the algorithms interpret dependencies.

F1 scores, my data shows around 0.133, which indicates some level of accuracy but not high compared to the manual scores. The manual shows a more variable range of F1 scores suggesting a better performance.

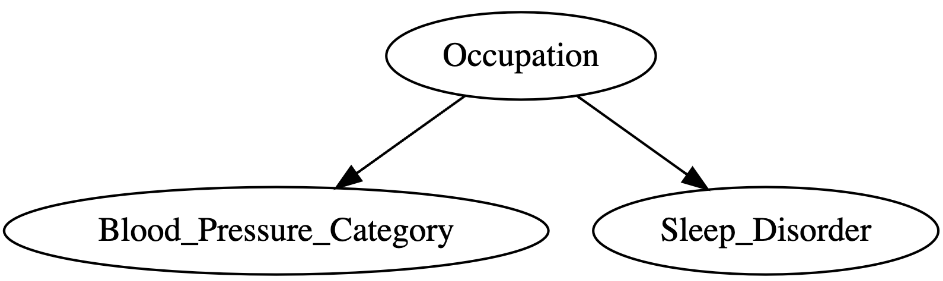
The reason for these differences can be various factors, my dataset sample size in less around 370 values. This difference can be linked to the complex things in real-world relationships in my DAGtrue, which may not have been analyzed by the algorithms.



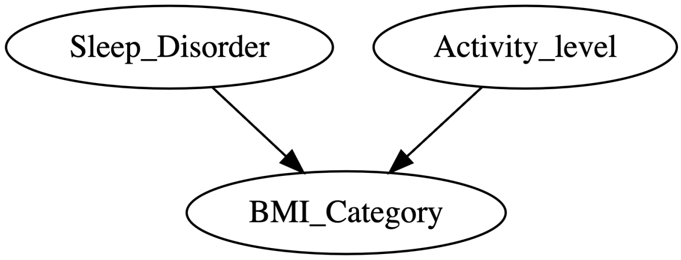
Causal Chain are the variables which have a direct effect on another variable, an example would be physical activity category to occupation via age group. Here an individual in physically active then his age group might be different and have different occupation.



Common Cause are the variables that have different effects, for example occupation might influence their blood pressure category, gender and sleep disorder. While a person can have similar occupation but can have different issues, such as blood pressure and sleep disorder.



Common effect are variables that cause a similar impact, an instance sleep disorder and activity level are both variables that can affect the BMI category. An individual with sleep disorder and less active may see a direct impact on their BMI category.



Rank	My scores			Ranking according to the Bayes manual		
	BSF	SHD	F1	BSF [average score]	SHD [average score]	F1[average score]
1	[GES] -0.082	[MAHC] 24	[MAHC] 0.138	[SaiyanH] 0.559	[MAHC] 50.96	[SaiyanH] 0.628
2	[MAHC] -0.082	[HC] 24	[GES] 0.133	[GES] 0.506	[SaiyanH] 57.98	[MAHC] 0.579
3	[TABU] -0.082	[TABU] 24	[TABU] 0.133	[MAHC] 0.503	[HC] 62.36	[GES] 0.552
4	[HC] -0.082	[GES] 24	[HC] 0.133	[TABU] 0.499	[TABU] 62.63	[TABU] 0.549
5	[SaiyanH] -0.161	[SaiyanH] 28	[SaiyanH] 0.118	[HC] 0.498	[GES] 63.2	[HC] 0.548

The analysis showed that the ranking of the algorithms differs from the manual. The MAHC algorithm ranks the highest in terms of the F1 score, suggesting it's best at identifying true relationships within the data. This partially aligns with the manual's findings. However, SHD for HC, TABU, GES, and MAHC is uniform, indicating a similar degree of accuracy among these algorithms in capturing the true data structure.

The SaiyanH algorithm, which performed well according to the manual, ranks lowest across all metrics in my dataset. This unexpected result highlights nature of algorithm performance, in which the complexity of underlying relationships can all have a significant impact on outcomes. The performance gap highlights the need for empirical testing on specific datasets to determine the efficacy of these algorithms. Since we used some real-world knowledge to make our DAGtrue, and with limited sample size of just 370+ the algorithm might not have identified the structure of the data resulting in poor ranking when compared to the manual.

Q6.

Algorithm	Structure learning elapsed time	Structure learning elapsed time manual
HC	0	10
TABU	0	13
SaiyanH	0	152
MAHC	0	272
GES	0	90

The results are consistent with those presented in the manual when compared to the dataset I used. This consistency is observed because, for all algorithms in the manual with a sample size of 10^2, the structure learning elapsed time is recorded as 0, which matches the runtimes I recorded for a similar sample size. But it doesn’t match with the average time of all sample sizes. Such outcomes can be attributed to factors like hardware capabilities, sample sizes, and the complexity of the dataset's structure.

Algorithm	Task 4 results			Algorithm	Task 5 results		
	BIC Score	Log Likelihood	Free parameters		BIC Score	Log Likelihood	Free parameters
Your knowledge-based graph	-7331.22	-4762.878	601	HC	-4496.936	-3291.824	282
				TABU	-4496.936	-3291.824	282
				SaiyanH	-4445.832	-3206.532	290
				MAHC	-4547.064	-3589.812	224
				GES	-4496.936	-3291.824	282

The knowledge-based graph gave a BIC score of -7331.22, which is significantly lower than the scores from the structure learning tasks, indicating that the model fits the data better due to its lower BIC value. The likelihood is also less negative (-4762.878), implying a better fit than the other results. However, it has 601 free parameters, which is more than double the parameters in the other models. This high number suggests a more complex model, which can be a sign of overfitting.

The BIC score penalizes the number of free parameters to strike a balance between data fit and model complexity. Despite its complexity, the lower BIC score indicates that the knowledge may have captured many true relationships that would otherwise be difficult to detect. The domain knowledge used determines whether these results meet expectations. If the knowledge is accurate and detailed, a more complex graph is possible.

In summary, the knowledge-based graph shows a better data fit but with greater complexity. This highlights the need for careful consideration of domain expertise in modeling to ensure a balance between a good fit to data and model parsimony.

Knowledge Approch	CPDAG Scores			LL	BIC	Free Parameters	Number of edges	RunTime
	BSF	SHD	F1					
Without Knowledge	-0.082	24	0.133	-3291.824	-4496.936	282	0	0
Directed	0.843	8	0.798	-3188.968	-6317.131	732	15	0
Undirected	0.275	17.5	0.394	-3187.208	-4755.563	367	8	0

When implementing of the knowledge constraints in the structure learning process, we observed some significant changes in the HC algorithm output. Particularly in the LL, BIC and F1 scores. In the initial case where there was no prior knowledge "Without knowledge" case represents the algorithms performance prior to any knowledge. Whereas directed and undirected cases the algorithms were fed with some domain knowledge integration.

Directed Approach: Here, we fed the algorithm with some causal directions like Physical activity -> Sleep Quality, which clearly tells that Physical activity influences the Sleep quality. This knowledge improved the F1 score from 0.133 to 0.798. This signifies a model's ability to identify a true causal relationship boosting both precision and recall. The LL also decreased suggesting a better fit of the model data. However, the BIC score increased from -4496 to -6317 which tells a potential overfitting or complexity introduced by the direct knowledge.

Undirected Approach: Here we added constraints like Age_Group - Physical_activity_Category offer a relation without specifying the direction of influence. This approach produced a F1 score of 0.394, which is higher than the baseline but lower than directed approach. This model outperforms the baseline in terms of accuracy while preventing overfitting. The smaller negative shift in the BIC score provides support for this claim.

The result aligns with the real-world relationships expectations. The directed knowledge was anticipated to give best model performance, but dint expect to overfit. Similarly, the undirected approach had small gains in BIC and F1 scores in line with predictions and balance between accuracy and model simplicity.

The shift in scores pre and post knowledge integration are quite different. The direct approach's high BSF and F1 scores indicate its best model performance within the dataset but also points to overfitting due to high BIC values. The undirect approach gives a balanced model with improved scores that suggest added knowledge without complexity.

constraintsDirected

ID	Parent	Child
1	Gender	Heart_Rate_Category
2	Gender	BMI_Category
3	Age_Group	Blood_Pressure_Category
4	Age_Group	Sleep_Disorder
5	Occupation	Stress_Level_Category
6	Sleep_Duration_Category	Sleep_Quality
7	Physical_Activity_Category	Activity_Level
8	Physical_Activity_Category	Heart_Rate_Category
9	Stress_Level_Category	Sleep_Quality
10	Stress_Level_Category	Blood_Pressure_Category
11	BMI_Category	Blood_Pressure_Category
12	Blood_Pressure_Category	Sleep_Disorder
13	Heart_Rate_Category	Activity_Level
14	Activity_Level	Sleep_Quality
15	Sleep_Quality	Sleep_Disorder

constraintsUndirected

ID	Parent	Child
1	Age_Group	Blood_Pressure_Category
2	Physical_Activity_Category	Heart_Rate_Category
3	Stress_Level_Category	Sleep_Quality
4	BMI_Category	Blood_Pressure_Category
5	Blood_Pressure_Category	Stress_Level_Category
6	Sleep_Duration_Category	Sleep_Quality
7	Occupation	Stress_Level_Category
8	Sleep_Quality	Sleep_Disorder

Reference:

[1] https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset/data
[2] https://ieeexplore.ieee.org/document/10199012

[3]https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10503965/#:~:text=Scientific%20literature%20shows%20that%20adults,15%2C19%2C20%5D.
[4] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4700549/
[5] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7723148/

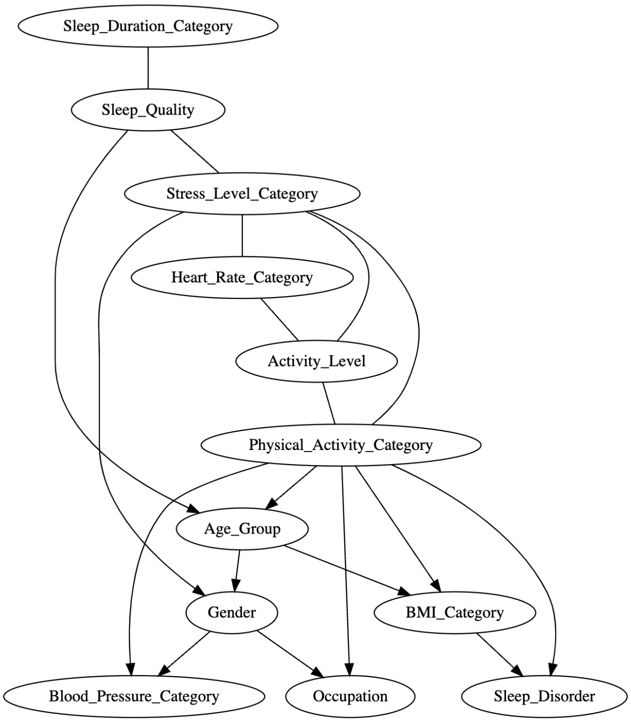
Appendix:

Task 3:
__________________________ CPDAG metrics __________________________
Precision score [CPDAG]: 1.000
Recall score [CPDAG]: 1.000
F1 score [CPDAG]: 1.000
SHD score [CPDAG]: 0.000
DDM score [CPDAG]: 1.000
BSF score [CPDAG]: 1.000
# of independent graphical fragments: 1 (includes 0 single-state variables)
____________________ Inference-based evaluation ____________________
LL for graph [log2]: -4762.878
BIC score [log2] -7331.220
# of free parameters 601

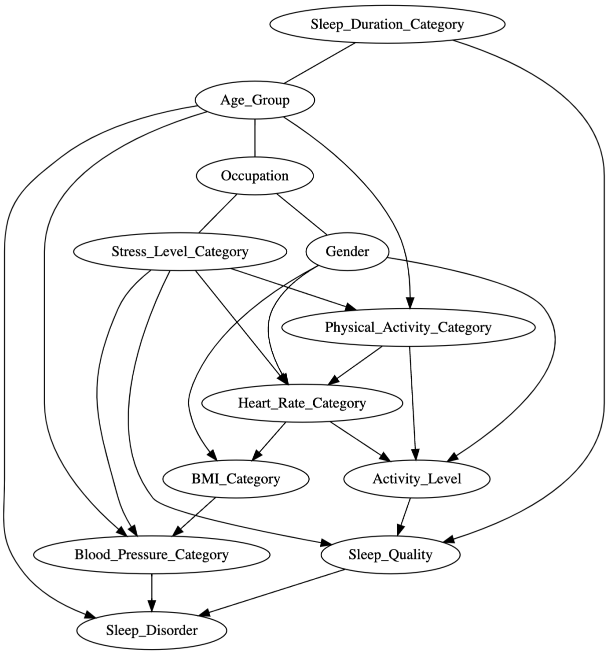
Directed Approach:
__________________________ CPDAG metrics __________________________
Precision score [CPDAG]: 0.652
Recall score [CPDAG]: 1.000
F1 score [CPDAG]: 0.789
SHD score [CPDAG]: 8.000
DDM score [CPDAG]: 0.467
BSF score [CPDAG]: 0.843
# of independent graphical fragments: 1 (includes 0 single-state variables)
____________________ Inference-based evaluation ____________________
LL for graph [log2]: -3188.968
BIC score [log2] -6317.131
# of free parameters 732
BUILD SUCCESSFUL (total time: 7 seconds)

Undirected approach:
__________________________ CPDAG metrics __________________________
Precision score [CPDAG]: 0.361
Recall score [CPDAG]: 0.433
F1 score [CPDAG]: 0.394
SHD score [CPDAG]: 16.500
DDM score [CPDAG]: -0.667
BSF score [CPDAG]: 0.276
# of independent graphical fragments: 1 (includes 0 single-state variables)
____________________ Inference-based evaluation ____________________
LL for graph [log2]: -3218.032
BIC score [log2] -4871.856
# of free parameters 387
BUILD SUCCESSFUL (total time: 6 seconds)

SaiyanH CPDAG graph:



Directed Approach:



Undirected Approach: