Question 2)

1. The joint distribution can be written as:

   Answer:
   $P(T,L,H,W,K,A,V,S)=P(L)\cdot P(H)\cdot P(W)\cdot P(K)\cdot P(T|L)\cdot P(V|W)\cdot P(S|L,H)\cdot P(A|S,T,W,K,V)$
   Equation represents the joint probability distribution of all the variables mentioned in the network.

2. Find the minimum number of parameters required to fully specify the distribution according to the above network.

   Answer:

   P(L): 1 parameter

   $P(H)$: 1 parameter

   $P(W)$: 3 parameters

   $P(K)$: 1 parameter

   $P(T|L)$: 4 parameters

   $P(V|W)$: 8 parameters

   $P(S|L,H)$: 8 parameters

   $P(A|S,T,W,K,V)$: 648 parameters

   Total number of parameters = 1 + 1 + 3 + 1 + 4 + 8 + 8 + 648 = 674

3. Answer the following questions

a) Write down a joint probability density function if there are no independence among the variables is assumed.
$P(T,L,H,W,K,A,V,S)=P(T)\cdot P(L|T)\cdot P(H|T,L)\cdot P(W|T,L,H)\cdot P(K|T,L,H,W)$
$\cdot P(A|T,L,H,W,K)\cdot P(V|T,L,H,W,K,A)\cdot P(S|T,L,H,W,K,A,V)$

b) Number of Parameters Required Without Independence Assumptions

Answer: The number of parameters required for each conditional probability table (CPT) is calculated as follows:

1. $P(T)$
   Parameters: 3−1=2
2. P(L|T)

   Parameters: 3× (2−1) =3

3. P (H|T, L)
   Parameters: 3×2x (2−1) =6

4. P(W|T, L,H)
   Parameters: 3×2×2 x (4-1) = 36

5. $P(K|T,L,H,W)$

   Parameters = 3×2×2 x4x (2-1) = 48
6. P(A|T,L,H,W,K)
   Parameters= 288
7. P(V|T,L,H,W,K,A):
   Parameters = 864
8. P(S|T,L,H,W,K,A,V)
   Parameters = 2592

   Total number of parameters = 2+3+6+36+48+288+864+2592=3839

C) Comparison and Comment

Answer:

Parameters with Independence Assumptions:

Total number of parameters: 674

Parameters without Independence Assumptions:

Total number of parameters: 3839

Explanation of the different parameters, assuming any independence among the variables (3839) is significantly higher than the number of parameters required when the independence assumptions specified by the Bayesian network are taken into consideration (674). This will depict the power of Bayesian networks in modelling complex systems.

2.4) From a previous study, the company, ShipComp, has found out that the ***Kilowatt Power (K) is conditionally independent of Length (L) given the Tonnage (T).*** The company ShipComp wants to modify the Bayesian network given in Figure 1 by incorporating this new information. **Assume** now that ***Kilowatt Power (K) is conditionally independent of Length (L) given the Tonnage (T)*****,** perform the following.

Answer:

Original Structure:

S depends on L and H

A depends on S,T,W,K,V

T depends on L

V depends on W

K depends on T

Modified Structure:

S depends on L and H

A depends on S,T,W,K,V

T depends on L

V depends on W

K depends on T

Re-draw the network.

Remove the edge $L \rightarrow K$

Add an edge $T \rightarrow K$

(b) Change in the Minimum Number of Parameters

Original Network Parameters (from Q2.2):

$P(L)$: 1

$P(H)$: 1

$P(W)$: 3

$P(K)$: 1

$P(T|L)$: 4

$P(V|W)$: 8 parameters

$P(S|L,H)$: 8 parameters

$P(A|S,T,W,K,V)$: 648 parameters

Total will be 674 parameters

New Network Parameters with conditional independence

$P(L)$: 1

$P(H)$: 1

$P(W)$: 3

$P(T|L)$: 4

$P(V|W)$: 8 parameters

$P(S|L,H)$: 8 parameters

$P(\mathrm{K}|\mathrm{T})$: 3 parameters

$P(A|S,T,W,K,V)$: 648 parameters

Total = 676 parameters

Comment on the results.
the change in the conditional independence structure will modify the dependency relationships, which will result in-to a slight increase in the complexity of the model.

2.5) d-separation method can be used to find two sets of independent or conditionally independent variables in a Bayesian network. Use the Bayesian network given in Figure 1 to answer the following:

a) Is Human factors (H) conditionally independent of Weather (W) given Kilowatt power (K), Severity of accident (S) and Accident type (A)?

Is **Human factors (H)** conditionally independent of **Weather (W)** given **Kilowatt power (K), Severity of accident (S)** and **Accident type (A)**?
Blocking/Non-blocking Analysis:

Path 1: $H \rightarrow S \rightarrow A \leftarrow W$

Conclusion: Non-blocking

Path 2: $H \rightarrow S \rightarrow A \leftarrow V \leftarrow W$

Conclusion: Blocking

Path 3: $H \rightarrow S \rightarrow A \leftarrow K \leftarrow T \leftarrow W$

Conclusion: Blocking

Path 4: $H \rightarrow S \leftarrow L \leftarrow T \leftarrow W$
Conclusion: Blocking

Path 5: $H \rightarrow S \leftarrow L \rightarrow T \leftarrow W$

Conclusion: Blocking

Final Conclusion: Path 1 is not -blocking while for the other path are blocking.

b. Is $\boldsymbol{T \perp V}$?

From the above analysis we observed that Path 1 is non-blocking. Therefore, T (Tonnage) and V (Visibility) are not independent in the mentioned Bayesian network. Hence, $T \not\perp V$.

2.6 For the Bayesian network shown in **Figure 1**, find all the nodes that are conditionally independent of **T (Tonnage**) given **A (Accident Type) and V (Visibility).**

Answer:

Paths from T to Other Network Nodes:

$T$to L(Length)

Path: $T \leftarrow L$

Blocked by conditioning on A and V: No

Conclusion: Not conditionally independent

T to H (Human factors)

Path 1: $T \rightarrow A \leftarrow S \leftarrow$H

Blocked by conditioning on A and V: Yes

Conclusion: Conditionally independent

T to S

Path 1: $T \rightarrow A \leftarrow S$

Blocked by conditioning on A and V: No

Conclusion: Not conditionally independent

T to W (Weather)

Path 1: $T \rightarrow A \leftarrow W$

Blocked by conditioning on A and V: Yes

Conclusion: Conditionally independent

T to K (Kilowatt Power)

Path: $T \rightarrow K$

Blocked by conditioning on A and V: No

Conclusion: Not conditionally independent

T to A (Accident Type)

A is one of the conditionings variables.

Conclusion: Not applicable

T to V (Visibility)

V is one of the conditioning variables

Conclusion: Not applicable

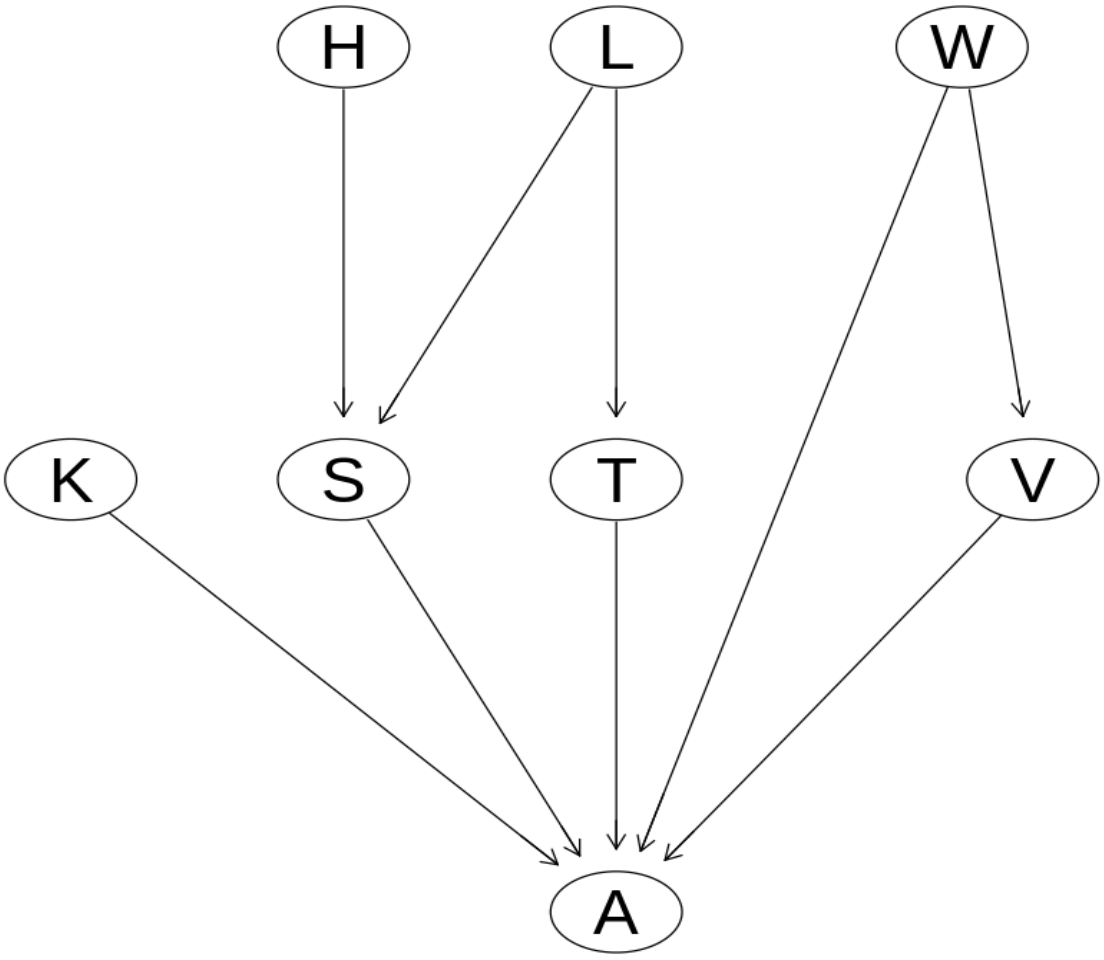
Conclusion:

Given A (Accident Type) and V (Visibility), the nodes that are conditionally independent of T (Tonnage) are:

H (Human factors)

W(Weather)

2.7) R code in code file



2.8) For the Bayesian network shown in **Figure 1**,

a. find the Markov blanket of A (Accident type).
Answer: Markov Blanket of A:

Parents of A: S,T,W,K,V
Children of A: None (in this structure)
Parents of Children of A: Since A has no children, this part is empty.
Therefore, the Markov blanket of A is: $\{S,T,W,K,V\}$

b. find all the nodes that are conditionally independent of A (Accident type) given its Markov blanket.

Answer: Nodes Conditionally Independent of Given Its Markov Blanket

Nodes in the Network:

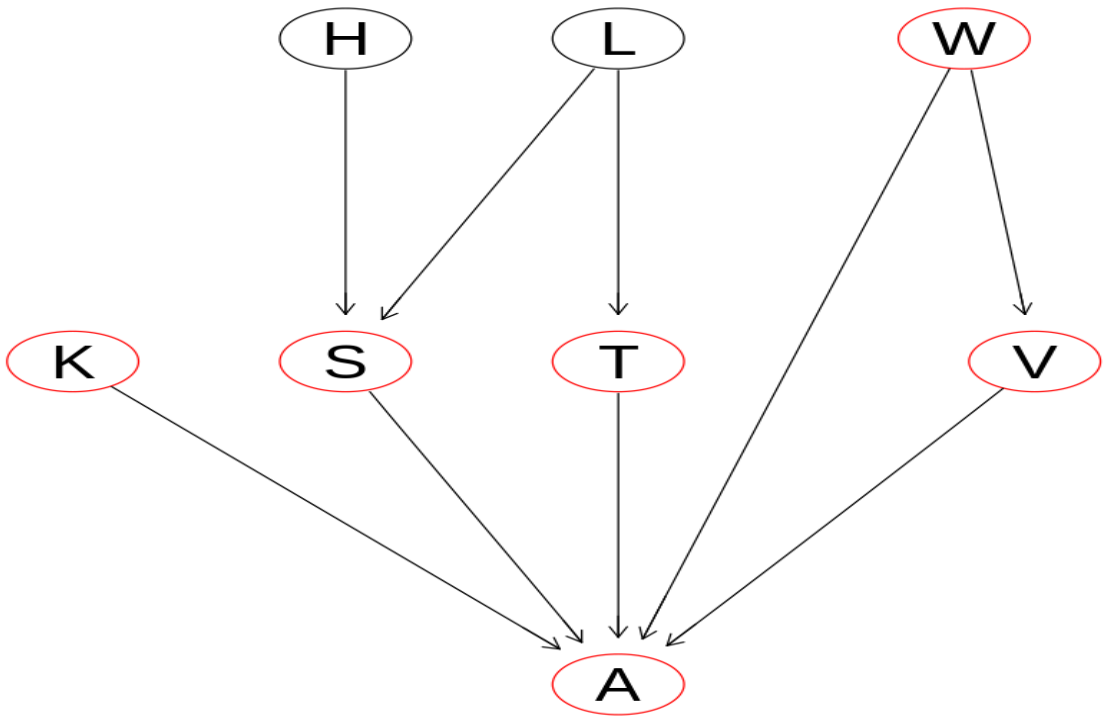
L: Length

H: Human factors

These nodes are not part of A's Markov blanket and do not have direct connections to A.

Therefore, nodes conditionally independent of A given its Markov blanket are: $\{L,H\}$

c) use R program to find the Markov blanket of K (Kilowatt Power). Plot the Bayesian network and show the Markov blanket nodes in the network using different colour.



2.9 For the Bayesian network shown in:

A) Step-by-Step Process for Variable Elimination to Compute

Combine the remaining factors to compute the probability of S:

P(S) ∝L,H,T∑ P(L)P(H)P(T|L)P(S|L,H) P(A=Collision|S,T,W=Fog ,K=Less than 3000,V=High)

b) Size of the largest clique: 4, treewidth =3

Question 3:

3.1 In the above network, state why A is independent of B, that is, A⊥B

Answer: In the given belief network, A is independent of B because there is no direct or indirect path connecting A to B. In other words, there are no arrows or edges between these two variables in the given directed acyclic graph (DAG) representing the network.
A only has an outgoing arrow to C. while B also only has an outgoing arrow to C.
Since A and B do not directly influence each other and their only interaction is through their common node C, A and B are independent from each other. This can be formally expressed

as $A \perp B$. This implies that the state of A provides no information about the state of B and vice versa.

3.2 Find $\boldsymbol{PP}(\boldsymbol{E}=\mathbf{0} \mid \boldsymbol{A}=\mathbf{0}, \boldsymbol{B}=\mathbf{1})$. Obtain the answer in terms of $\boldsymbol{\beta}$ (give the answer in a simplified form). Show all the steps clearly.

Ans:

Step 1: Identify the Relevant Conditional Probabilities:

- From the graph we know that A and B influence C, while C will D and E.

Step 2: Find the Joint Distribution:

Need to calculate $P(A=0, B=1, C, E=0)$

Step 3: Marginalize over c:

Calculate $P(C \mid A=0, B=1)$

$$P(C \mid A=0, B=1) = \begin{cases} 0.5 & \text{if } C=0, \\ 0.2 & \text{if } C=1, \\ 0.3 & \text{if } C=2. \end{cases}$$

$$P(E=0 \mid C) = \begin{cases} 0.4 & \text{if } C=0, \\ \beta & \text{if } C=1, \\ 0.8 & \text{if } C=2. \end{cases}$$

Law of Total Probability

$P(E=0|A=0,B=1)=C\sum P(E=0|C)P(C|A=0,B=1)$

$P(E=0|A=0,B=1)=0.44+0.2\beta$

3.3 : The table shown below provides 30 simulated data obtained for the above Bayesian network.

Answer: Alpha =Number of instances where A=0/ Total Number of instance = 7/30≈0.233

β= Number of instances where C=1 and E=0/ Total number of instances where C=1

β= 5/10 = 0.5

3.4: Find the value of $\boldsymbol{PP(EE=00|\ AA=00,BB=11)}$ using the appropriate values obtained from the above question **Q3.3**.

P(E=0|A=0,B=1)=0.44+0.2β

P(E=0|A=0,B=1)=0.44+0.2×0.5

P(E=0|A=0,B=1)=0.54

3.5 In R code

Question 4) Whole covered in the Code Part

Question 5)

a. Variables Used and Algorithm for Learning the Bayesian Network Structure
   - Accident Severity: The severity of the accident.
   - Road Surface Condition: The condition of the road surface at the time of the accident.
   - Weather Condition: During accident weather conditions
   - Vehicle Type: The type of vehicle involved in the accident.
   - Time of Day: The time when the accident occurred.
   - Accident Type: The type of accident, such as a collision, rollover, etc.
   - Traffic Volume: The volume of traffic at the time of the accident.
   - Driver Experience: The experience level of the driver involved in the accident.

   The algorithm used for learning the Bayesian network structure in this study is the "Parents and Children" algorithm, which is a constraint-based approach.

b. Independence of Injury Type and Sex Given Education, Seat Belt, Licence Type, and Vehicle Type.

   Variable Elimination Process for $P(S|W$=Fog,A=Collision,V=High,K=Less than 3000)P(S|W=Fog,A=Collision,V=High,K=Less than 3000)

- Since there is a direct path from "Sex" to "Injury type" that is not blocked by any of the of the given variables ("Education," "Seat belt," "License type," "Vehicle type"), "Injury type" is not independent of "Sex" given the knowledge about these variables.

c. Explanation of Probabilities Shown for the Nodes in Figure 5.

Conditional Probabilities: For each node, the probability distribution is conditioned on the states of its parent nodes. For instance, the probability of "Injury Type" is conditioned on the values of its parent nodes, such as "Vehicle Type," "Seat Belt," and others as per the structure learned.

Example Nodes:

Injury Type: The conditional probabilities here would indicate the likelihood of various injury types (e.g., minor, severe) given specific conditions like vehicle type, seat belt usage, etc.

Vehicle Type: The conditional probabilities for this node show the likelihood of different types of vehicles involved in accidents, conditioned on relevant parent nodes.

These given probabilities help in understanding the influence of various factors on the likelihood of different outcomes and provide a probabilistic framework to analyse and predict accident characteristics based on observed data.

d. Parameter Learning in the Road Accident Network

i. The Probability of Being Not-Injured While Wearing a Seat Belt and Driving a Car, Knowing That the Driver Has a Diploma Degree and a Type 2 Driving License

P(Not-Injured|B=Wearing,V=Car,E=Diploma,L=Type 2)

P(Not-Injured|Wearing,Car,Diploma,Type 2)=0.70

ii. The Probability of Being Dead While Wearing a Seat Belt and Driving a Car, Knowing That the Driver Has a Diploma Degree and a Type 2 Driving License.

P(Dead|Wearing,Car,Diploma,Type 2)=0.05

Question 6)

Bayesian network analysis of road accident data using the provided dataset "Crash_Reporting_-_Drivers_Data.csv". The analysis of the dataset involves selecting relevant variables which include exploratory data analysis, constructing Bayesian networks using different structure learning algorithms, and comparing the resulting networks.

Selection of variables:

1. Injury.Severity (Categorical): Indicates the severity of the injury sustained in the accident.
2. Vehicle.Body.Type (Categorical): Describes the type of the vehicle involved in the accident.
3. Weather (Categorical): The education level of the driver (e.g., No formal education, Diploma, Degree).
4. Collision.Type (Categorical): The type of collision that occurred.
5. Surface.Condition (Categorical): The condition of the road surface at the time of the accident.
6. Light (Categorical): The lighting conditions during the accident.
7. Speed.Limit: The speed limit in the area where the accident occurred, converted into categorical levels.
8. Driver.Substance.Abuse (Continuous, Discretized): Age of the driver.

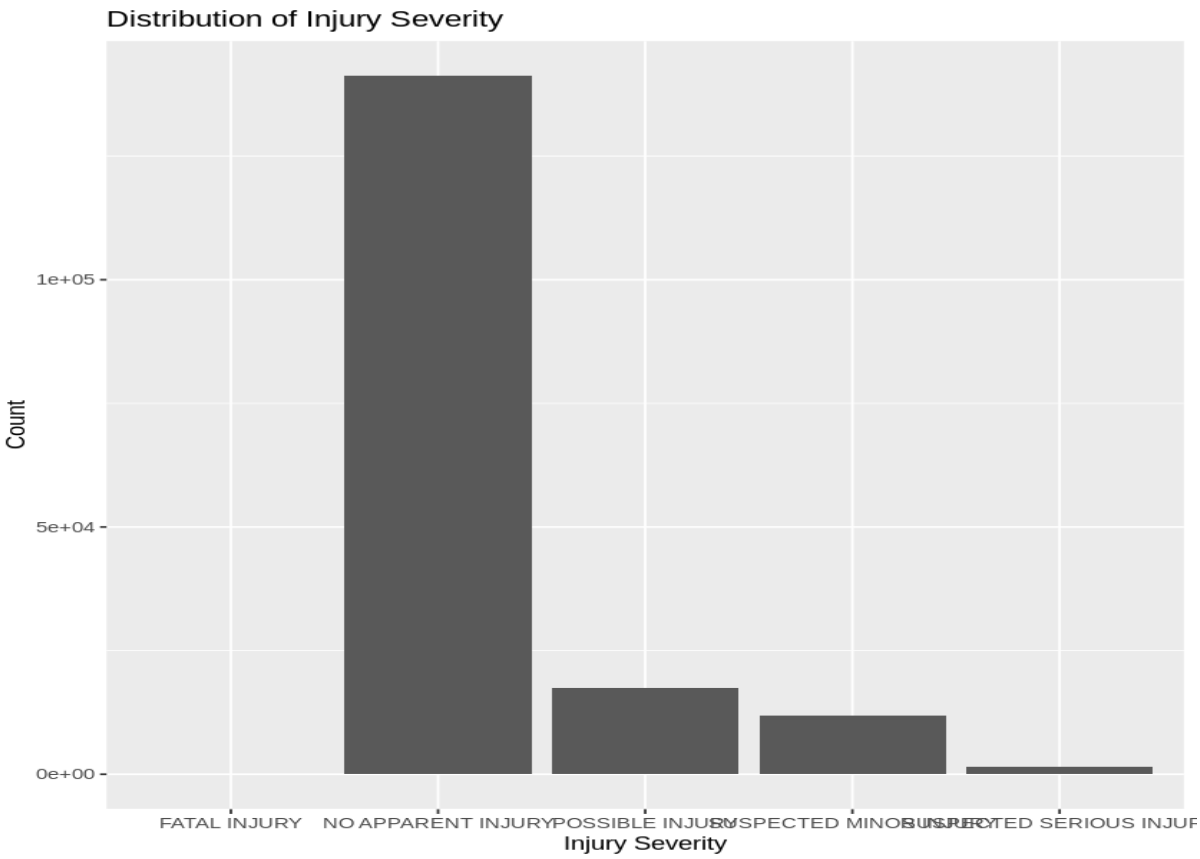
Using the R code, we can generate the summary statistics and visualizations provide insights into the distribution of variables. For example, the bar plots show the distribution of driver age, injury types, and weather conditions during accidents.

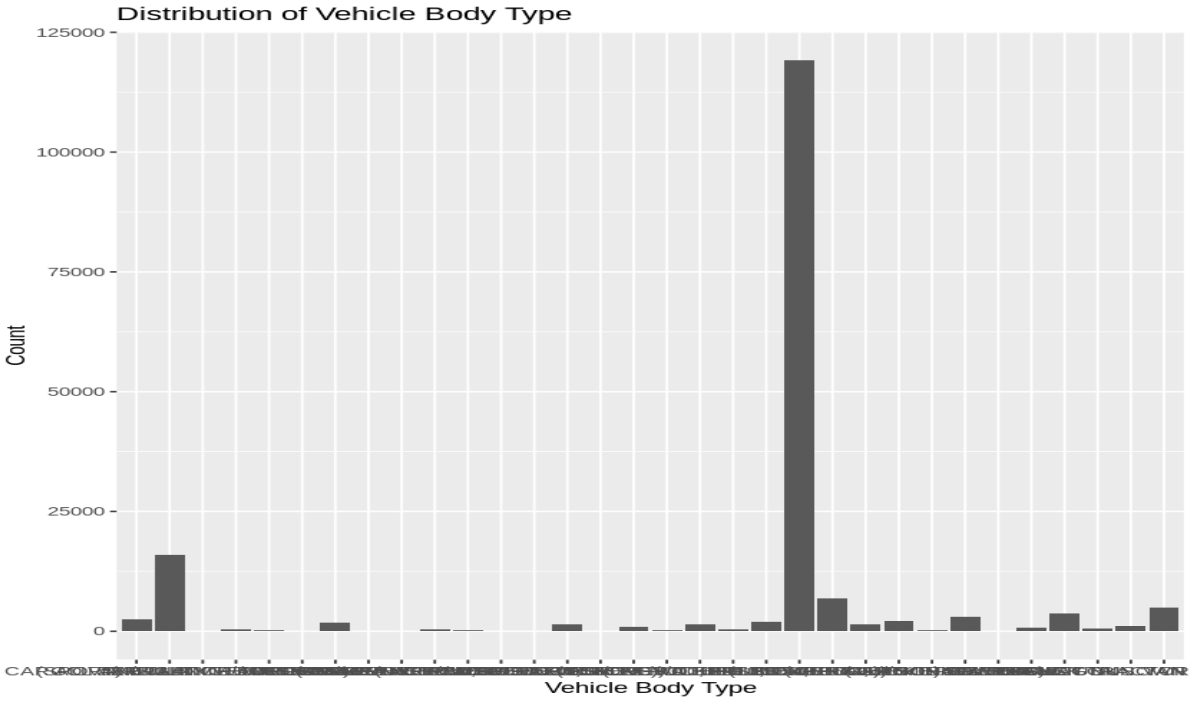
Comparison of Bayesian Networks

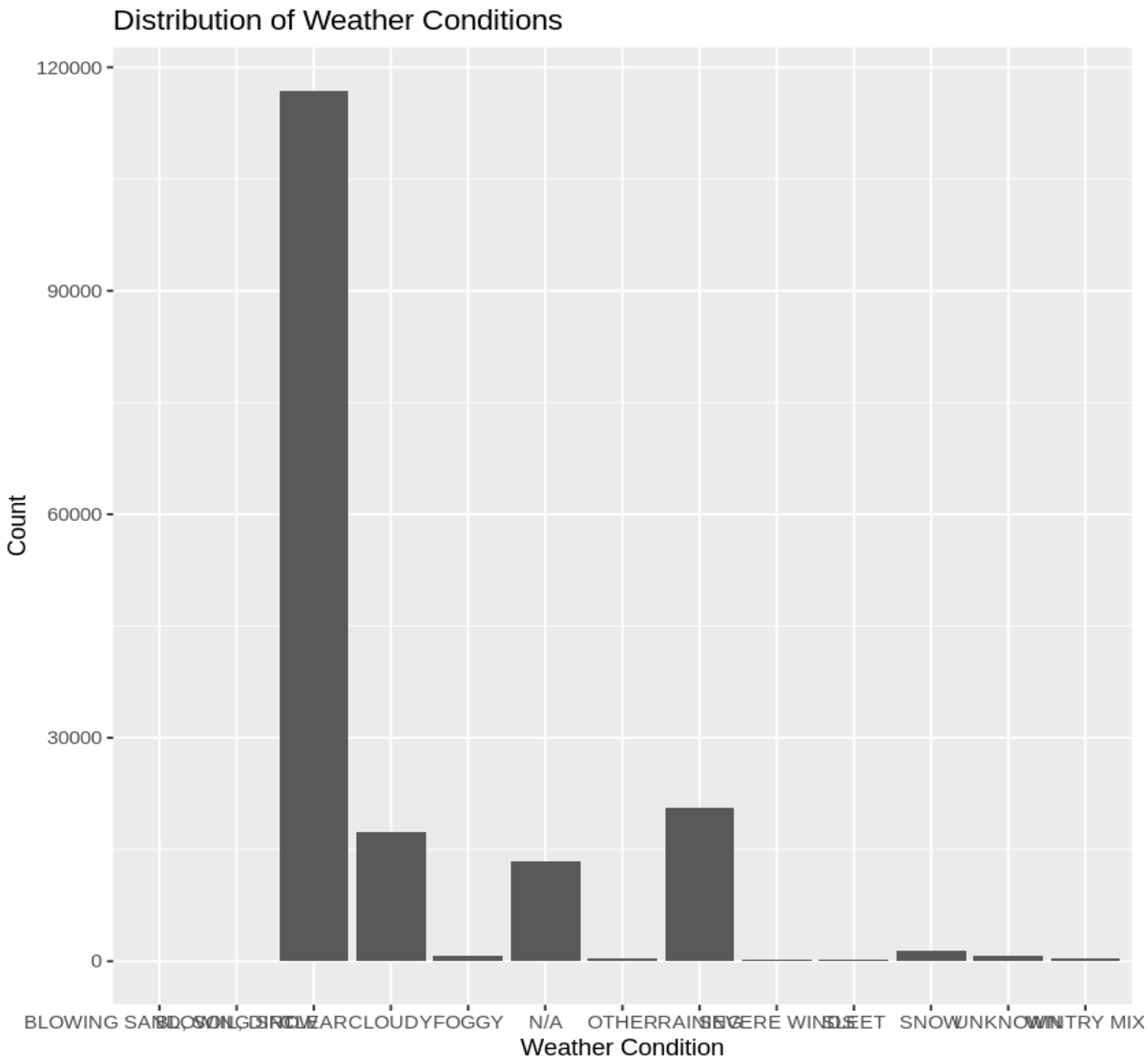
We compare the networks learned using different algorithms in terms of their structure and parameter estimates. The comparison helps in understanding the strengths and weaknesses of each method and selecting the most suitable one for our analysis.

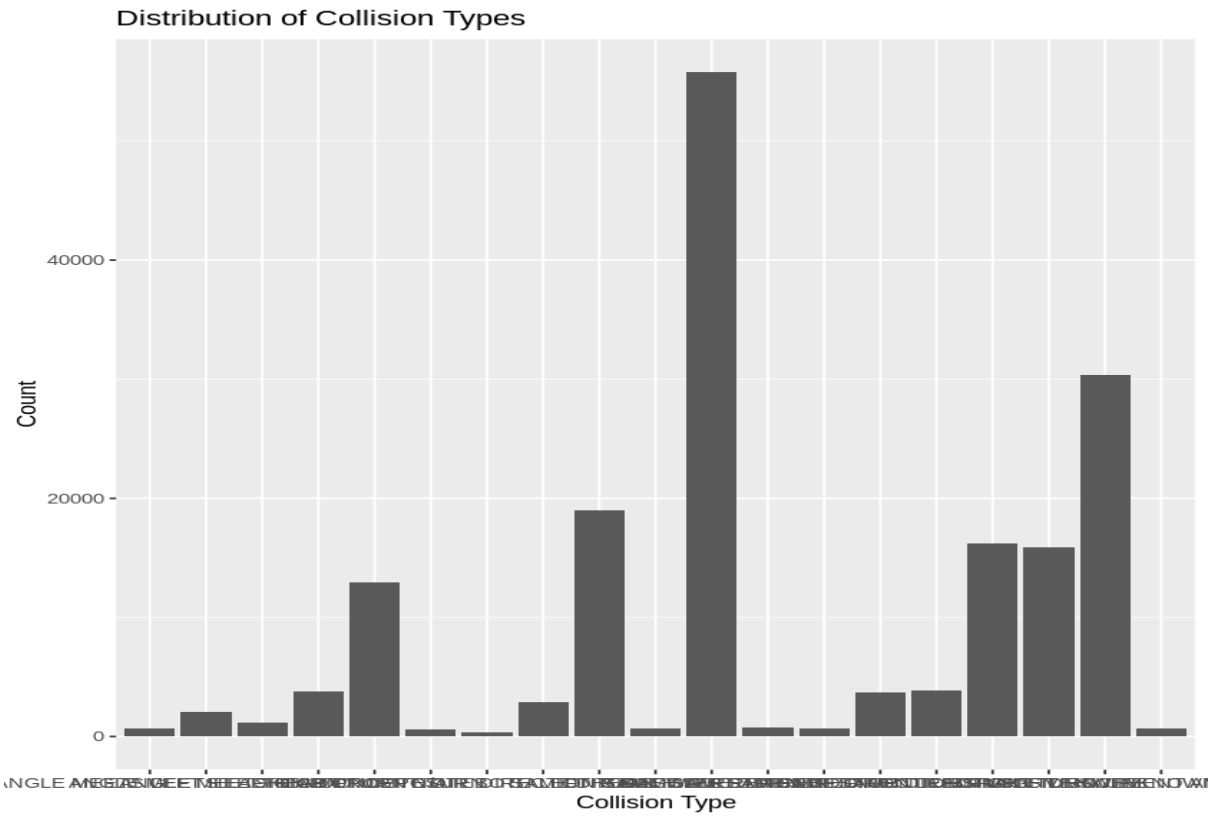
```
Injury.Severity    Vehicle.Body.Type    Weather          Collision.Type
 Length:172105      Length:172105      Length:172105      Length:172105
 Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character
```

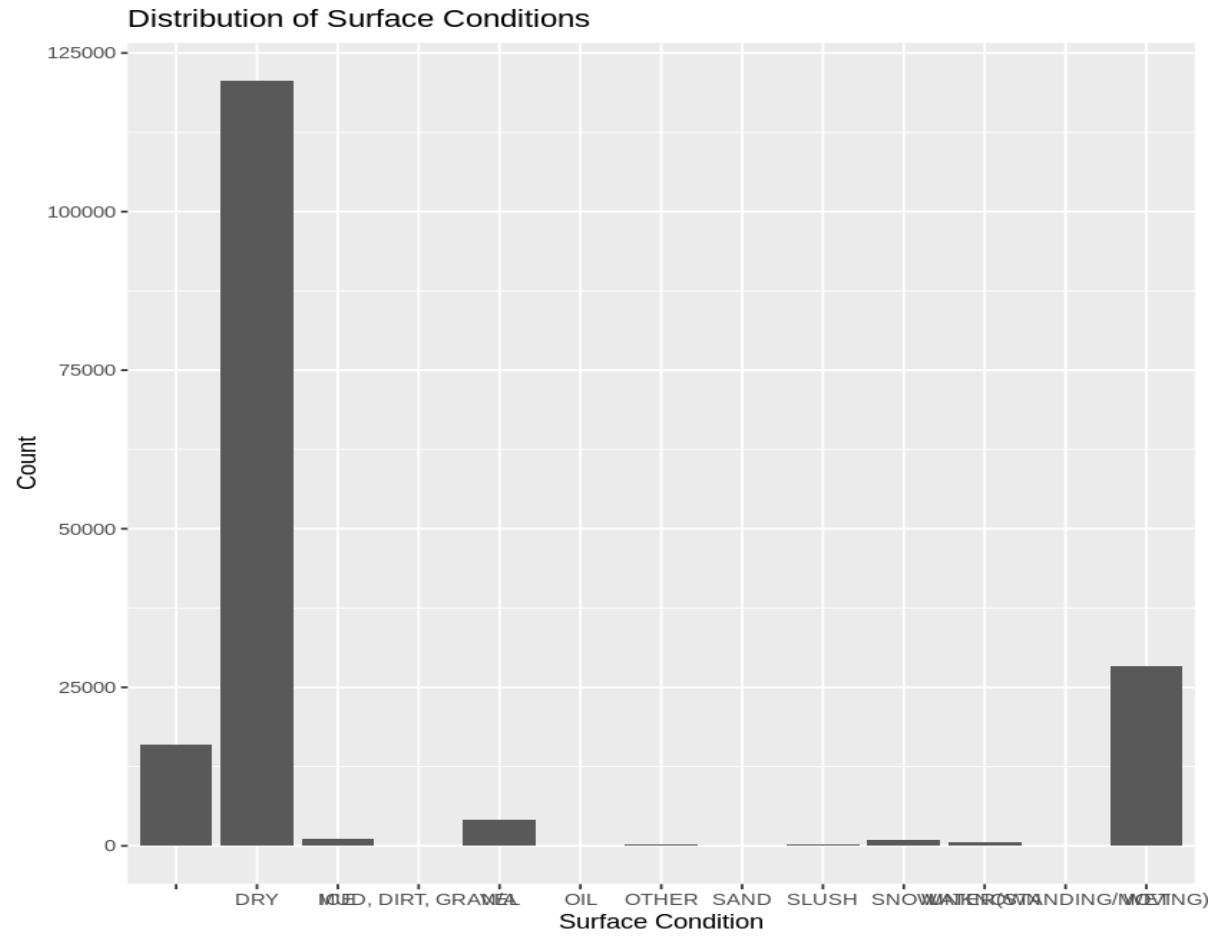
```
 Surface.Condition     Light          Speed.Limit     Driver.Substance.Abuse
 Length:172105      Length:172105      0-30  : 62046   Length:172105
 Class :character   Class :character   31-50 :101191   Class :character
 Mode  :character   Mode  :character   51-70 :  4118   Mode  :character
                                       71-90 :     1
                                       91-110:     0
                                       111+  :     0
                                       NA's  :  4749
```

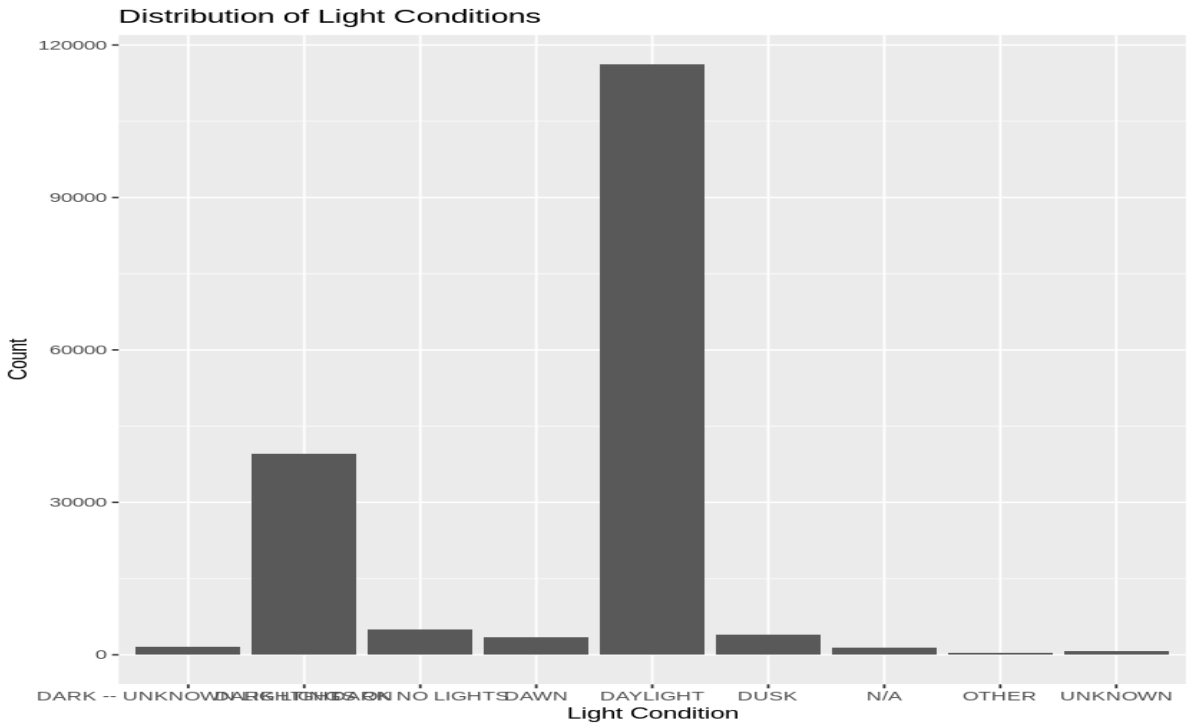

Distribution of Injury Severity
Count
1e+05
5e+04
0e+00
FATAL INJURY
NO APPARENT INJURY
POSSIBLE INJURY
SUSPECTED MINOR INJURY
SUSPECTED SERIOUS INJUR
Injury Severity

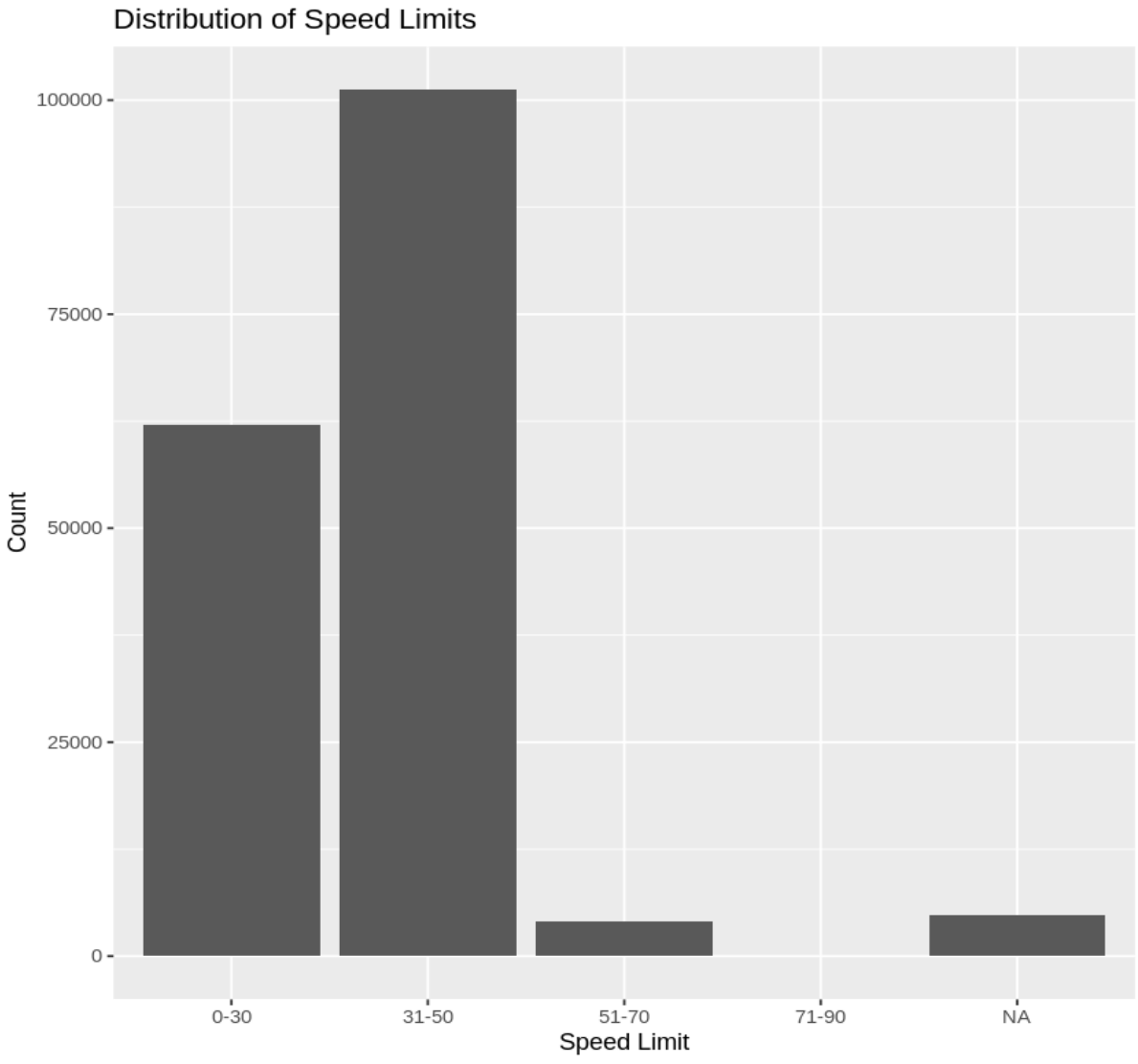

Distribution of Vehicle Body Type
Count
125000
100000
75000
50000
25000
0
Vehicle Body Type

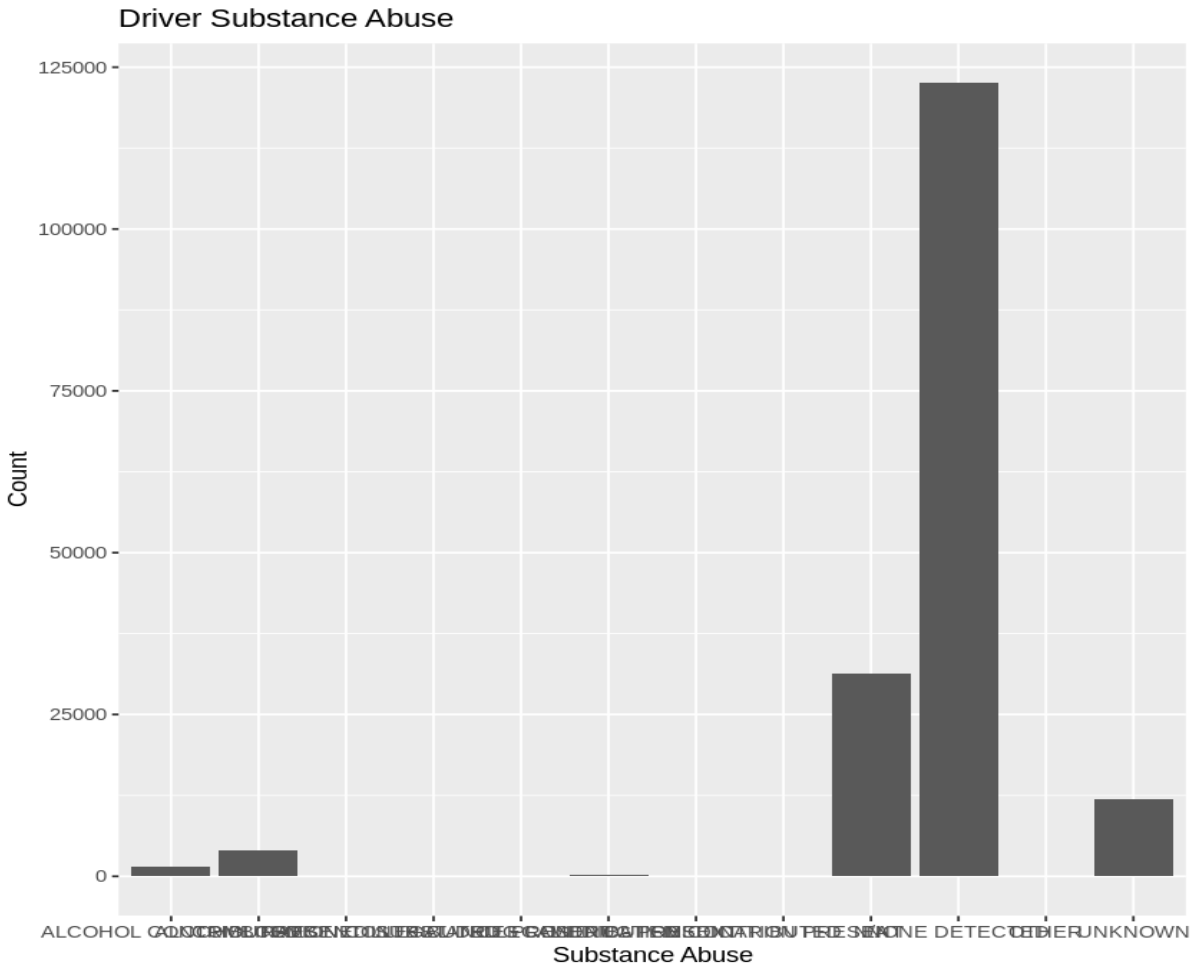

Distribution of Weather Conditions
120000
90000
60000
30000
0
Count
BLOWING SAND
CLEAR
CLOUDY
FOGGY
N/A
OTHER
RAIN
SEVERE WINDS
SLEET
SNOW
UNKNOWN
WINTRY MIX
Weather Condition


Distribution of Collision Types
Count
40000
20000
0
Collision Type

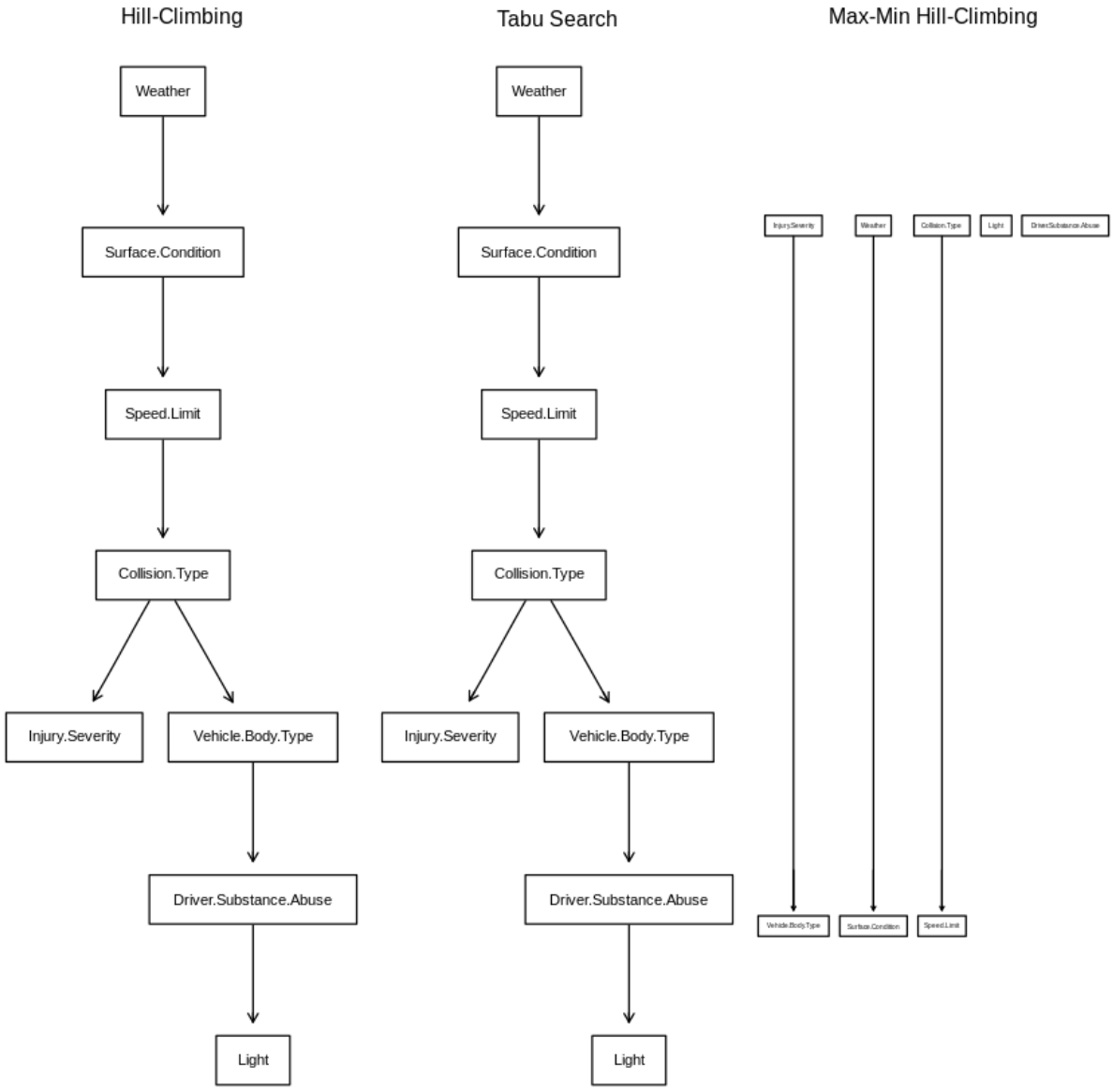

Distribution of Surface Conditions
Count
125000
100000
75000
50000
25000
0
DRY
OIL
OTHER
SAND
SLUSH
Surface Condition


Distribution of Light Conditions
120000
90000
60000
30000
0
Count
DARK -- UNKNOWN
NO LIGHTS
DAWN
DAYLIGHT
DUSK
N/A
OTHER
UNKNOWN
Light Condition


Distribution of Speed Limits
100000
75000
50000
25000
0
Count
0-30
31-50
51-70
71-90
NA
Speed Limit


Driver Substance Abuse
125000
100000
75000
50000
25000
0
Count
ALCOHOL
NONE DETECTED
OTHER
UNKNOWN
Substance Abuse

Bayes Network Formation



Model Testing:

P(No Apparent Injury | Vehicle.Body.Type=PASSENGER CAR, Driver.Substance.Abuse=NONE DETECTED) = 0
P(Dead | Vehicle.Body.Type=PASSENGER CAR, Driver.Substance.Abuse=NONE DETECTED) = 0

The results provide insights into the factors affecting road accidents and inform strategies for improving road safety.