Walkability Scores in Corpus Christi

Jacob Harris

Walkability Index Scores

Introduction

Published in a report from June 2021, the Environmental Protection Agency defines walkability as follows:

'The definition of walkability is simple: a walkable place is easy to walk around. Walkable communities come in various sizes and styles depending upon where they are located in the country; whether they are in a city, suburban area, or small town; and whether they have public transit. A walkable community in a small Northeastern village could look very different from a Southwestern city.' [1]

According to the EPA report, walkable cities are easier to navigate and increased exercise for individuals will also have added benefits for public health concerns such as obesity or diabetes rates. [2]

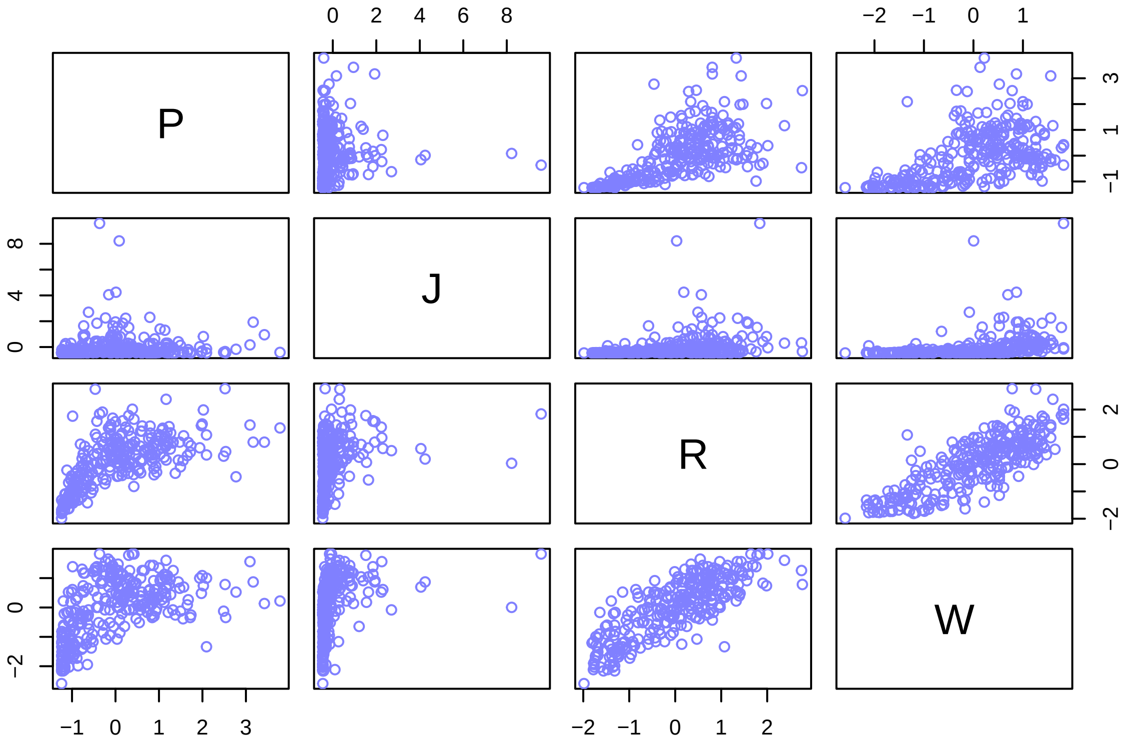
This project takes a look at influences of walkability for the city of Corpus Christi. The main concern is if the population density, density of jobs, and road network affect the walkability score for a location within the city.

Data

The data was originally published by the EPA in the June 2021 National Walkability Index report. This data is available to the public and can be found on data.gov [3]. The original report gives a score from 1-20 with 20 as most walkable for locations within the United States. The max walkability score for an area in Corpus was 18, with the lowest being 1.

This data set comes from the Smart Location Database (SLD), which combines Census datasets, HERE Maps NAVSTREETS data, US Geological Survey datasets, and transit location, service, and schedule data. Each record in the EPA report summarizes several variables for Census block groups. These block groups are a way to section a region or city [4]. Usually, these block groups are no larger than 600 to 3000 people. The SLD includes block groups for all 50 states and territories in the US, resulting in 117 different factors (some administrative in nature) for over 220,000 records.

To reduce the scope of this project into a more manageable task, census block groups located in the city of Corpus Christi were selected and only 3 variables were chosen to look at influence on walkability, resulting in a dataset with 287 records. The dataset used for this project looks as follows



The original variables are:

P - Population Density (People/Acre)

J - Job Density (Jobs/Acre)

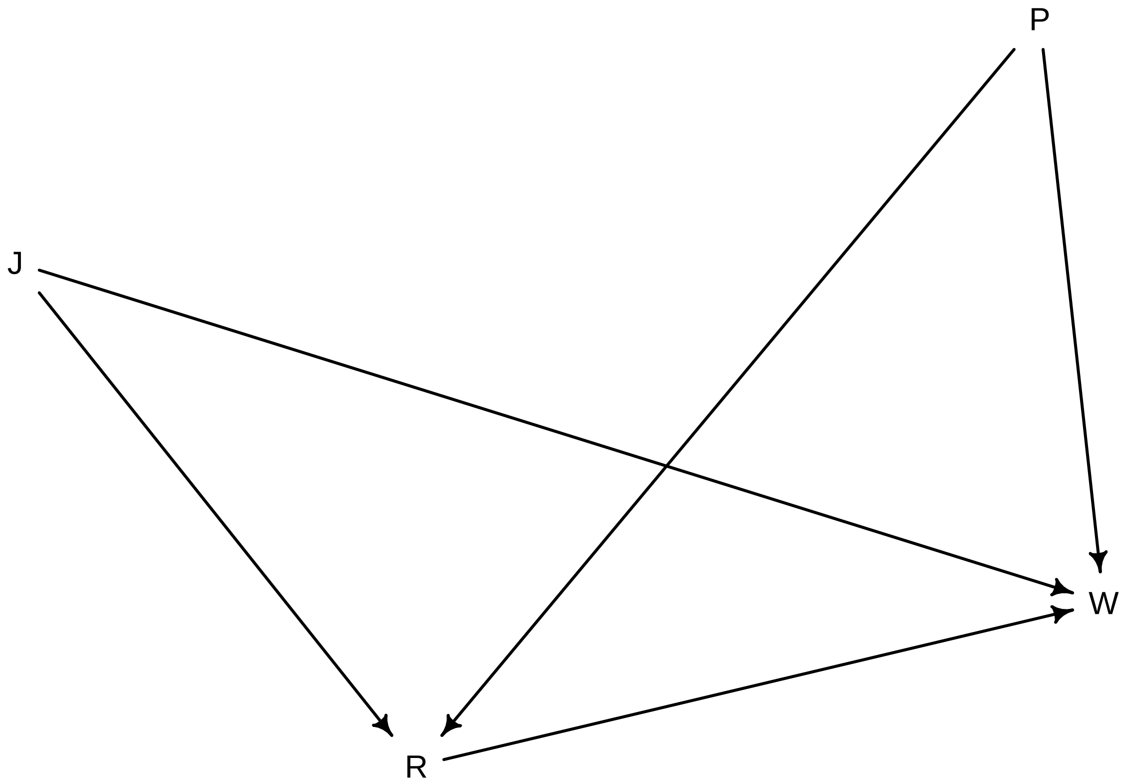
R - Total Road Network Density

W - Walkability Score (0-20)

All variables used in this project are standardized.

Directed Acyclic Graph

In order to look at the causal influences affecting walkability, we need to have a graph that is supported by the dataset. The following is a result of looking at the dataset and ruling out a proposed number of DAGs.



In this graph, we assume all three predictors of interest influence walkability. This graph also assumes road density is affected by the population and job density. Note that this graph assumes job and population density have no affect on each other. We can test this DAG by checking the conditional independencies assumed by the DAG, in this case job density is independent of population density

In order to check this, we can use a simple linear regression model to predict job density from population density in a census block group. The model is

$$J \sim N(\mu, \sigma)$$
$$\mu = \alpha + \beta_P * P$$
$$\alpha \sim N(0, 0.2)$$
$$\beta_P \sim N(0, 0.5)$$
$$\sigma \sim Exp(1)$$

```
##        mean    sd   5.5% 94.5%
## a     0.000 0.056 -0.090 0.090
## bp    0.075 0.058 -0.018 0.168
## sigma 0.994 0.041  0.928 1.060
```

The data supports the DAG above when checking the independence of these two varaibles. Looking at the mean (0.07) and standard deviation (0.06) of bp, population density alone provides no information in predicting job density. As a result, we can say that both are independent of each other.

Models

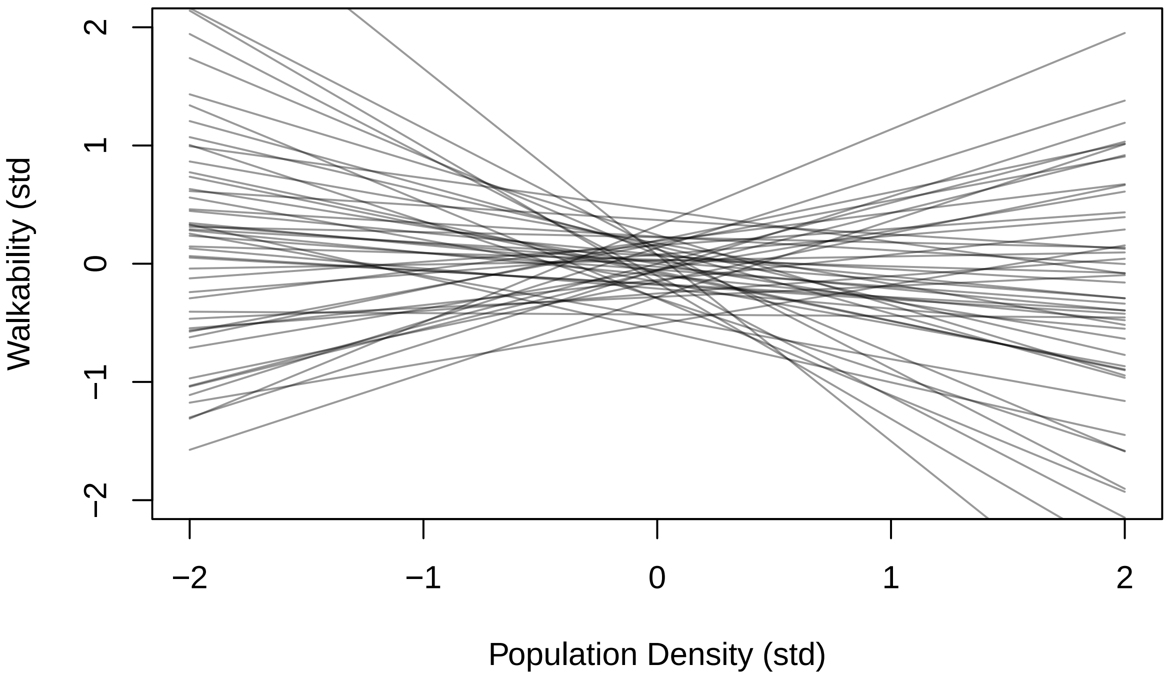
To check the influences of population, job, and road network density, we can check whether or not adjustment sets need to be accounted for depending on the predictor.

Population Density Predictor Model

To determine the effect of population density on walkability, we can create a model that uses just population density as a predictor of walkability. The model looks as follows.

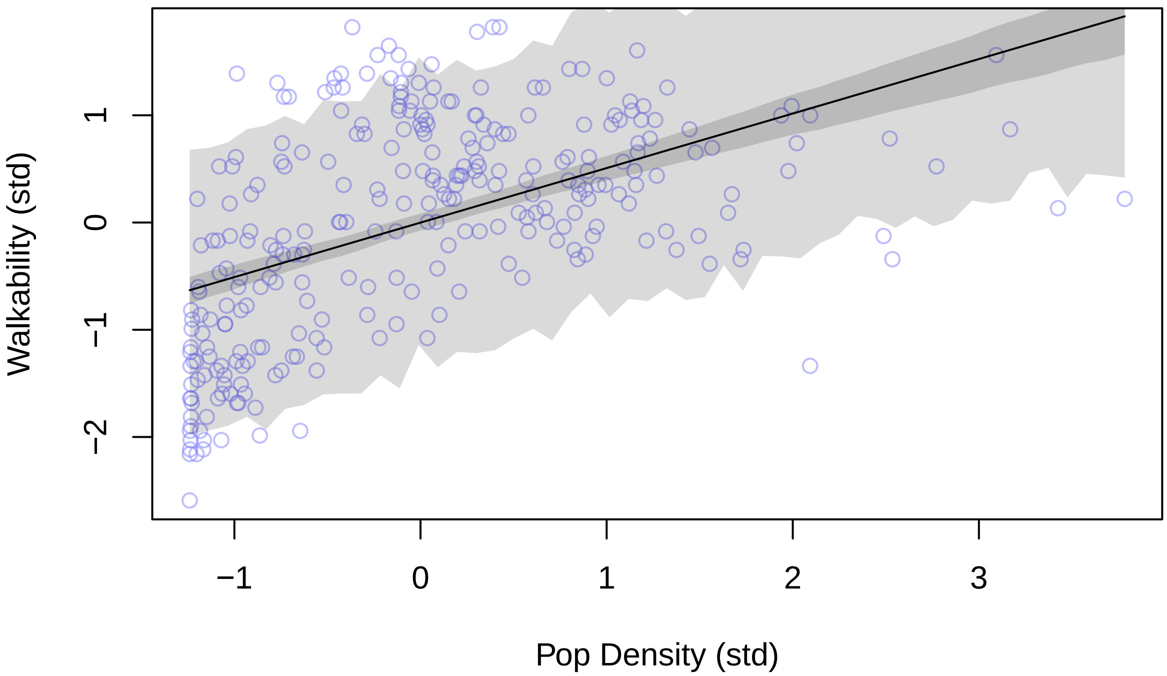
$$W \sim N(\mu, \sigma)$$
$$\mu = \alpha + \beta_P * P$$
$$\alpha \sim N(0, 0.2)$$
$$\beta_P \sim N(0, 0.5)$$
$$\sigma \sim Exp(1)$$

This graph shows a plot of possible lines sampled from the priors in this model. By showing no preference of sign for the slope and limiting extreme slope values, this can be considered as an uninformative prior for our model.




```
##         mean    sd   5.5% 94.5%
## a      0.000 0.049 -0.078 0.078
## bP     0.508 0.050  0.427 0.588
## sigma  0.856 0.036  0.799 0.913
```

For the population density model, an increase in 1 standard deviation of population density (5.35 people/acre) would indicate an increase in .51 standard deviations of walkability (an increase of 1.96). The model having a small standard deviation for the population density coefficient indicates a confidence in the estimate for bP.



The plot above shows the 89% prediction interval for walkability as a function of population density along with the average line for walkability at a given population density. All values are standardized. The narrow shaded region is the 89% interval for mu while the larger shaded region is where the model expects 89% of the values for standardized walkability at a given value.

Job Density Predictor Model

To determine the effect of job density on walkability, we can create a model that uses just population density as a predictor of walkability. The model looks as follows.

$$W \sim t(\mu, \sigma)$$

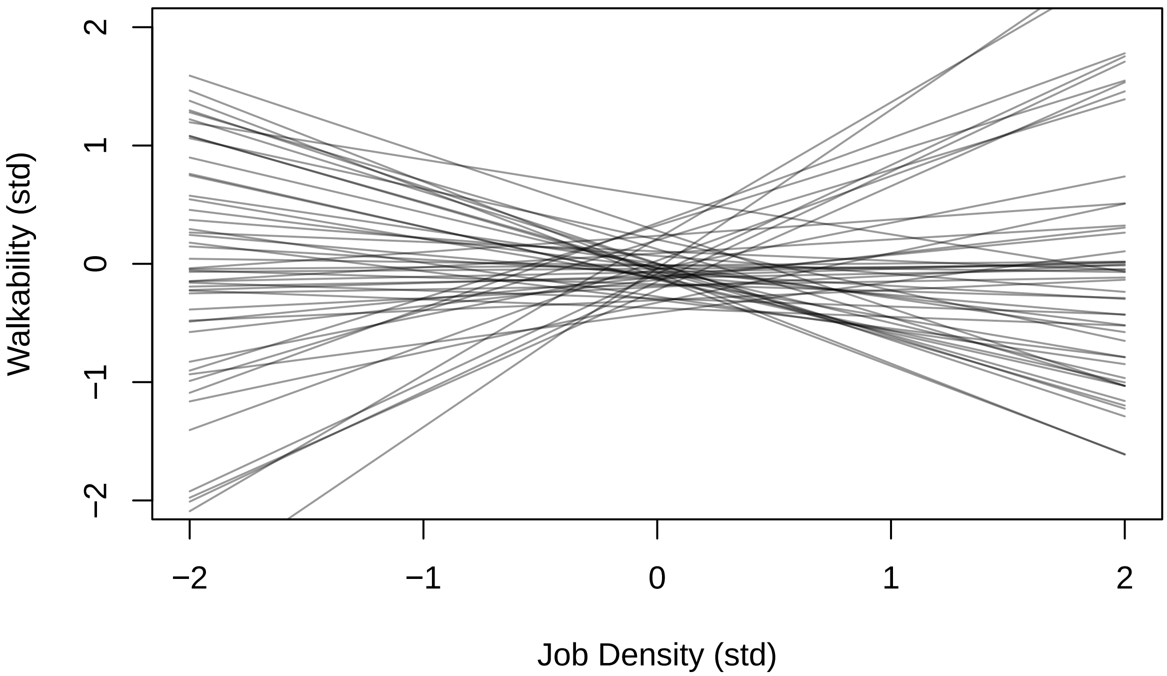
$$\mu = \alpha + \beta_J * J$$

$$\alpha \sim N(0, 0.2)$$

$$\beta_J \sim N(0, 0.5)$$

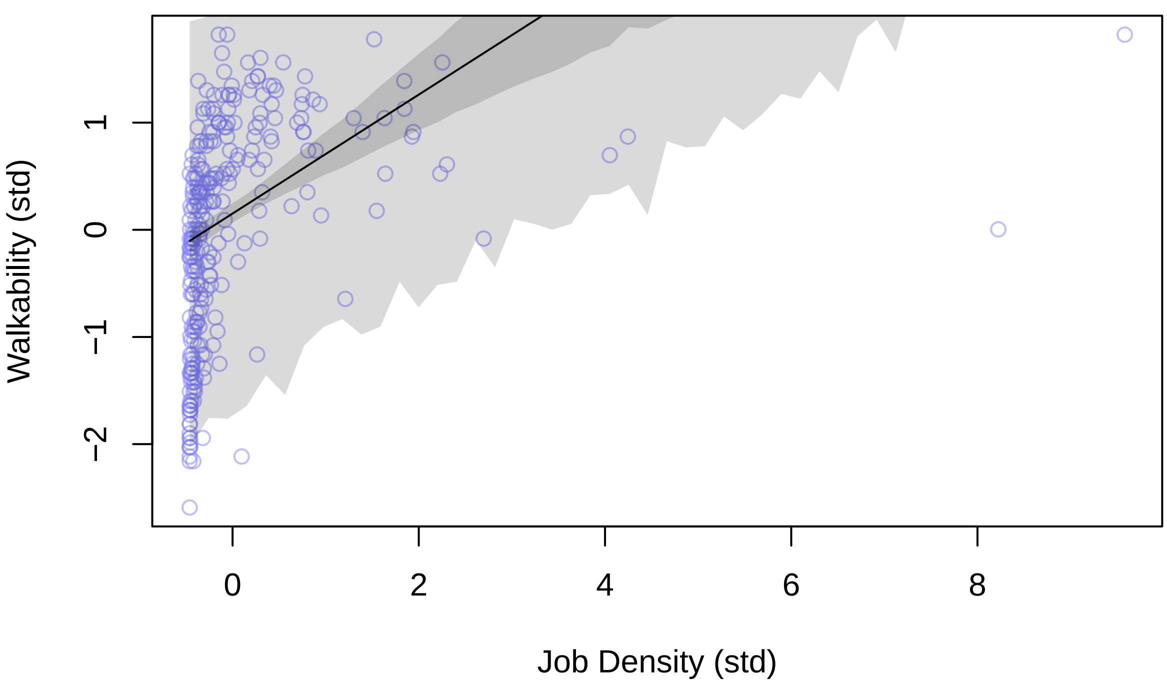
$$\sigma \sim Exp(1)$$

We can also look at the priors for this model as well. Similar to the population density predictor model, these are uninformative priors, as no preference for positive or negative association is given. This model uses a Student t distribution as the likelihood for walkability scores. When a Gaussian likelihood is used, highly influential points affect the posterior.




```
##         mean    sd  5.5% 94.5%
## a     0.152 0.056 0.063 0.241
## bJ    0.557 0.108 0.385 0.729
## sigma 0.700 0.044 0.631 0.770
```

The table above gives the summary for the model using job density as a predictor for walkability. For each 1 standard deviation increase in job density (4.38 jobs/acre), the model expects an increase of 0.56 standard deviations in walkability (an increase of 2.15).



The graph above shows the 89% prediction interval for walkability given a job density (both standardized). This graph shows the interval for mu in the smaller shaded region. This interval becomes more uncertain as the value of standardized job density increases.

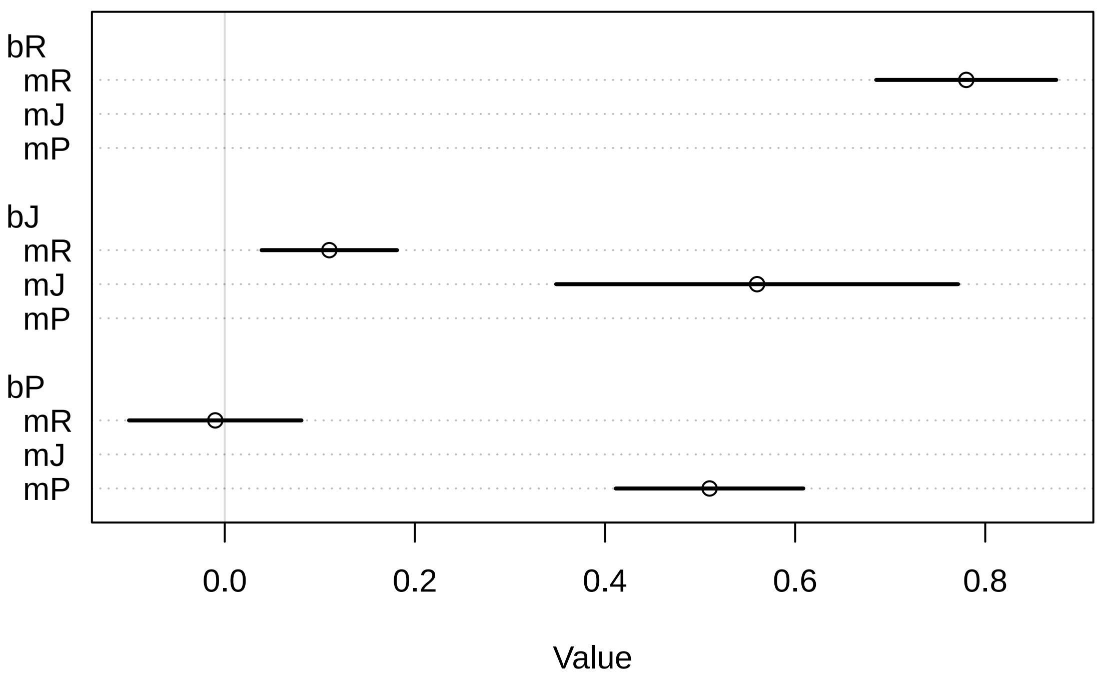
Road Density Predictor Model

In order to look at the causal influence of road density on walkability, we need to account for population density and job density. This is the model that conditions on those variables as well

$$W \sim N(\mu, \sigma)$$
$$\mu = \alpha + \beta_R * R + \beta_J * J + \beta_P * P$$
$$\alpha \sim N(0, 0.2)$$
$$\beta_R \sim N(0, 0.5)$$
$$\beta_J \sim N(0, 0.5)$$
$$\beta_P \sim N(0, 0.5)$$
$$\sigma \sim Exp(1)$$

```
##         mean    sd   5.5% 94.5%
## a      0.000 0.034 -0.054 0.054
## bR     0.778 0.048  0.701 0.856
## bJ     0.108 0.036  0.050 0.166
## bP    -0.012 0.046 -0.086 0.062
## sigma  0.580 0.024  0.542 0.619
```

The table above shows the summary of the model using road density as a predictor of walkability. When we account for job density and population density, the model shows road density is informative in explaining walkability for a given area. An increase of 1 standard deviation of road density (8.72) results in an increase of 0.78 standard deviations for walkability (an increase of 3.00).



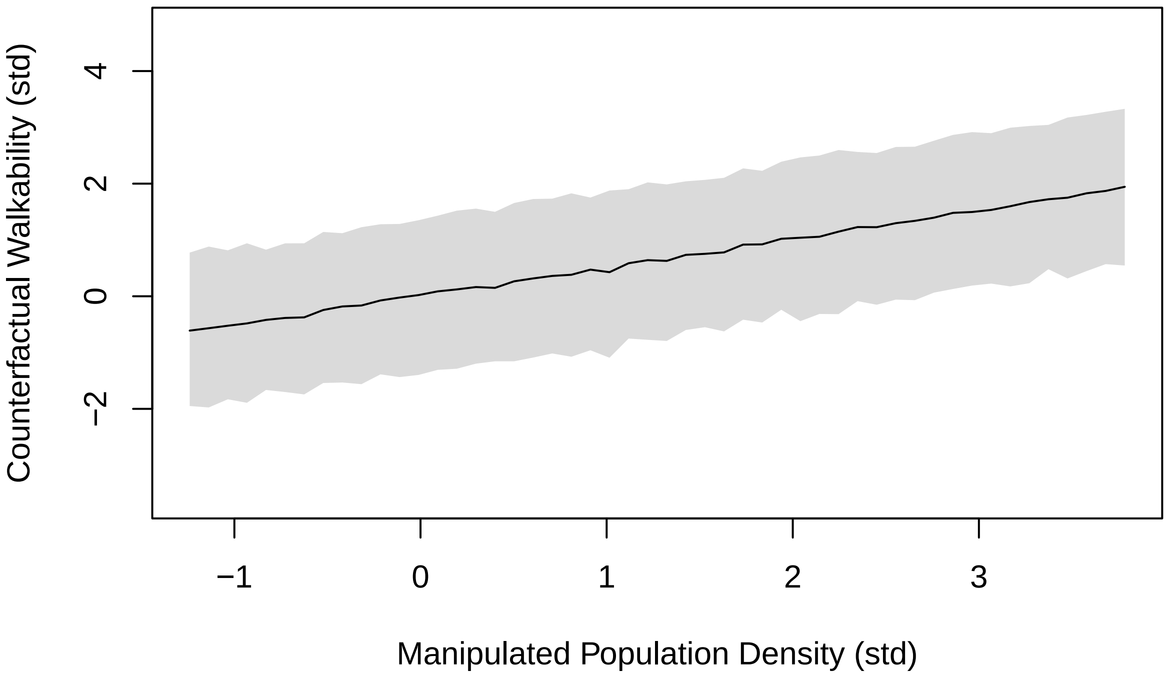
The graph above shows a comparison between the three models used to show causal influences of walkability. All models show that there is value in using population density (bP), job density (bJ), and road density (bR) in explaining the walkability of a given area. Each predictor expects an increase in walkability as that standardized predictor increases in value.

Counterfactual Plots

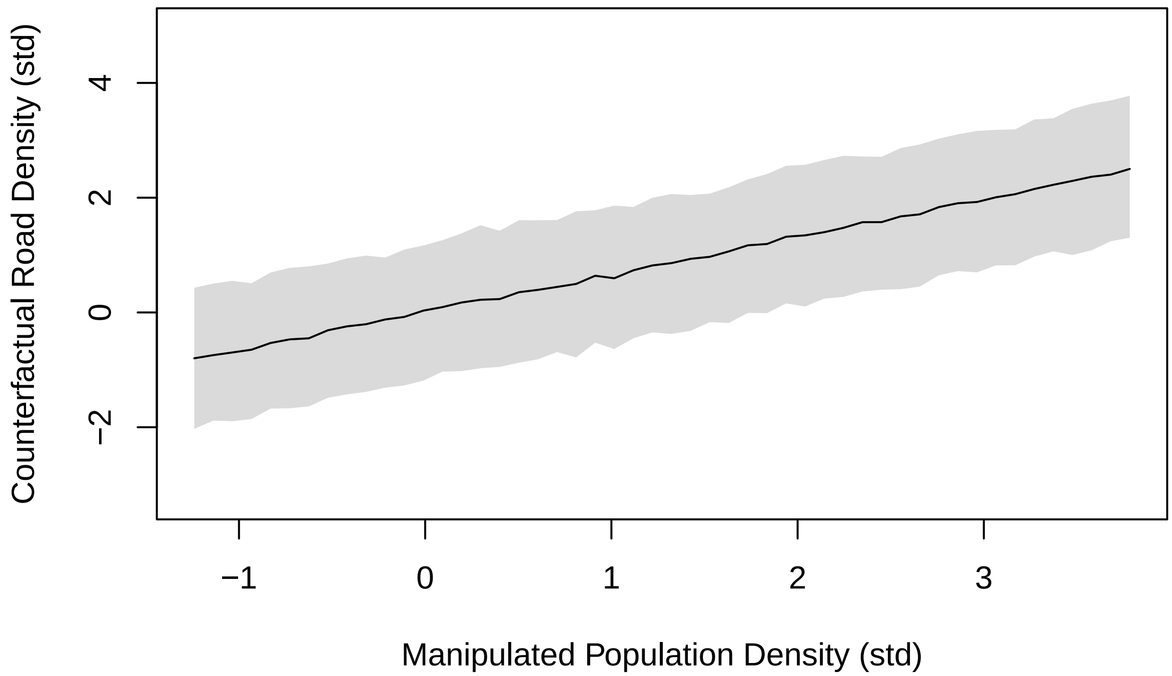
Using the DAG, we can look at how changing each predictor influences walkability and other predictors. First we can take a look at how population density affects road density and walkability

```
##            mean    sd   5.5% 94.5%
## a         0.000 0.034 -0.055 0.055
## bR        0.826 0.046  0.752 0.900
## bP       -0.035 0.046 -0.109 0.039
## sigma     0.589 0.025  0.550 0.628
## aPR       0.000 0.043 -0.069 0.069
## bPR       0.659 0.044  0.589 0.729
## sigmaPR   0.746 0.031  0.696 0.795
```

The table above shows that population density is positively associated with road density and we can simulate the effects of changing population density on walkability and road density.



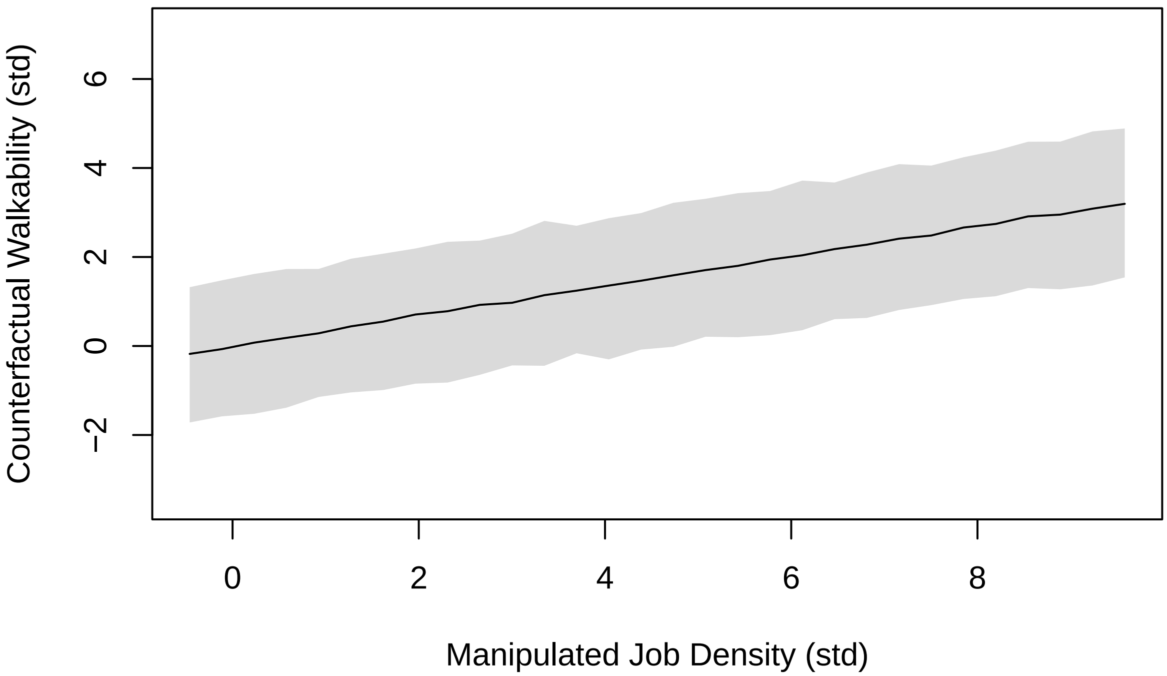
The graph above shows different manipulated values of standardized population density and the resulting values of walkability. As shown before, an increase in population density should result in an increase in walkability



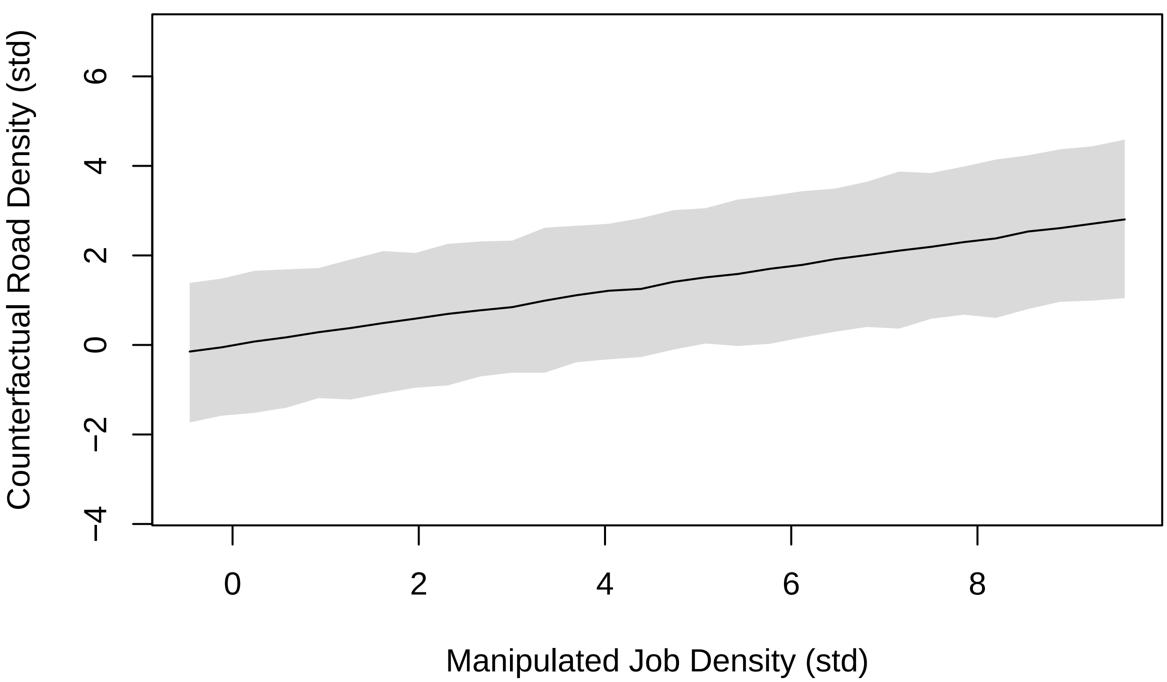
This graph above shows the effects of changing values of standardized population density on standardized road network density. The model shows a positive association, so increasing population would result in an increase in road density.

```
##          mean    sd   5.5% 94.5%
## a       0.000 0.034 -0.054 0.054
## bR      0.770 0.036  0.713 0.827
## bJ      0.110 0.036  0.053 0.167
## sigma   0.580 0.024  0.542 0.619
## aJR     0.000 0.054 -0.086 0.086
## bJR     0.294 0.056  0.205 0.384
## sigmaJR 0.951 0.040  0.888 1.015
```

The table above shows job density has a positive association with road density and we can look at the effects of changing job density.



The graph above shows the effects of changing standardized job density on standardized walkability. Since there is a positive correlation between the two, increasing job density results in an increase in walkability, however, the effect is less than that seen by increasing population density.

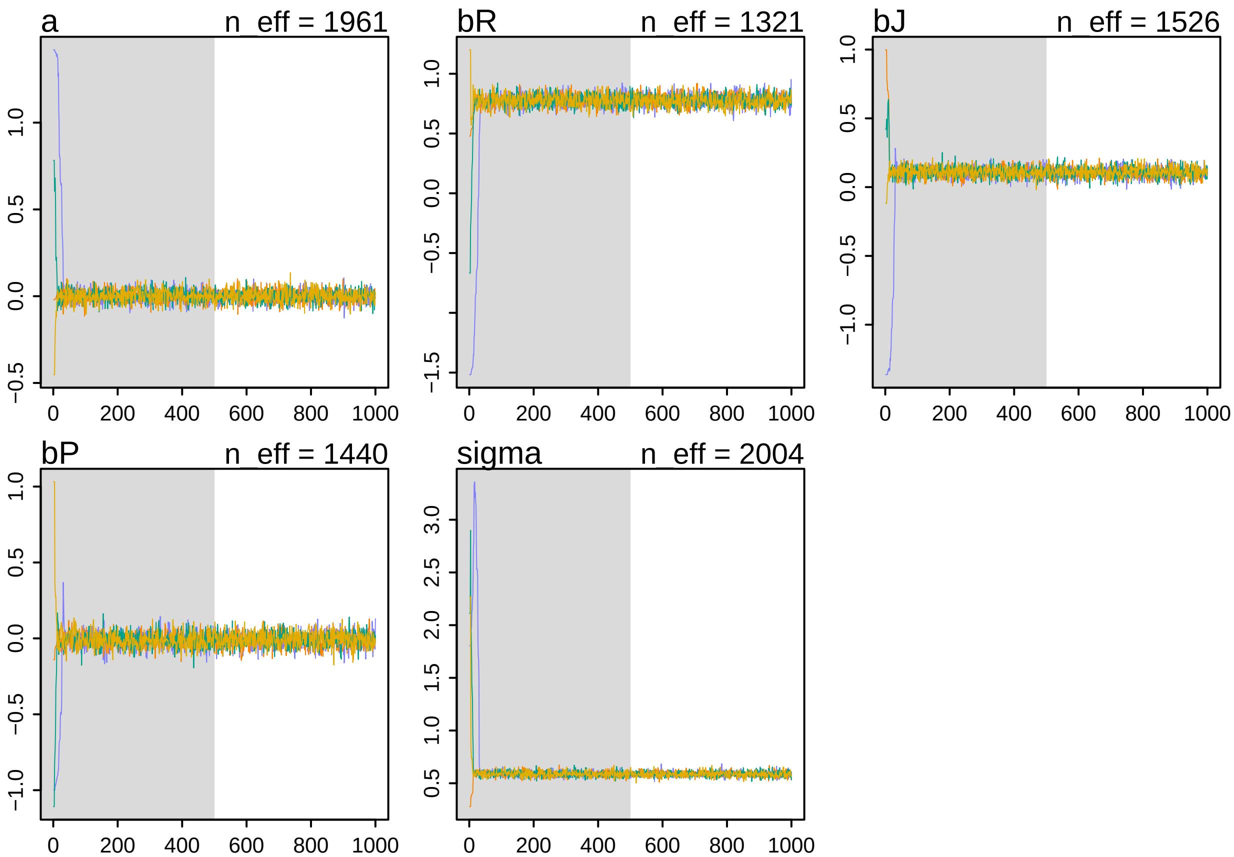


The graph above shows the effect of increasing job density on road density. An increase in standardized job density should result in an increase in standardized road density.

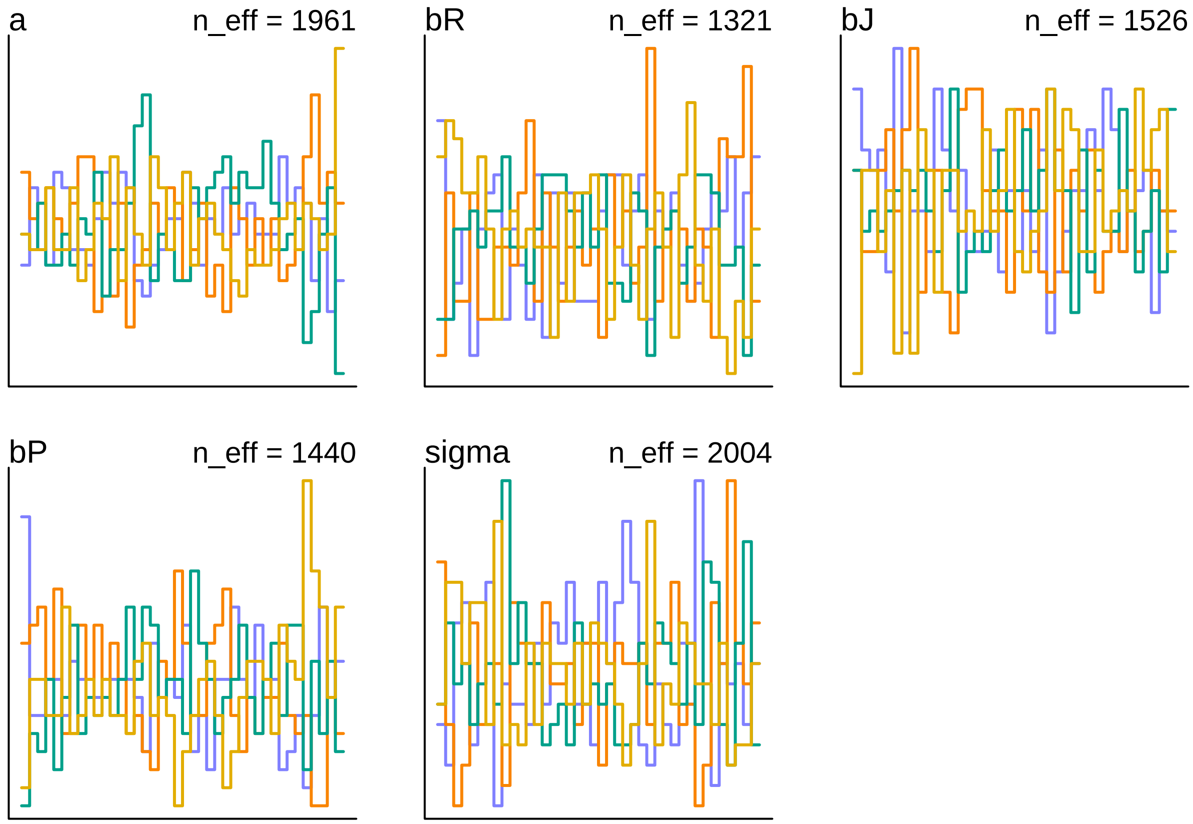
Model Assumptions

This project uses quadratic approximation instead of Hamiltonian Monte Carlo for models. Since the variables are standardized, the assumption that the posterior is relatively multinormal should hold. The model 'mJ' benefits from using a robust regression due to a few influential points in the data. Below is a comparison of the mR model using Hamiltonian Monte Carlo and quadratic approximation.

First we can check the chains for the Hamiltonian Monte Carlo. The traceplot and trankplot below shows the chains converging and having a good effective number of samples.



The charts above show the chains converging for the Hamiltonian Monte Carlo. The initial variance is expected as the chain start from different locations in sampling.



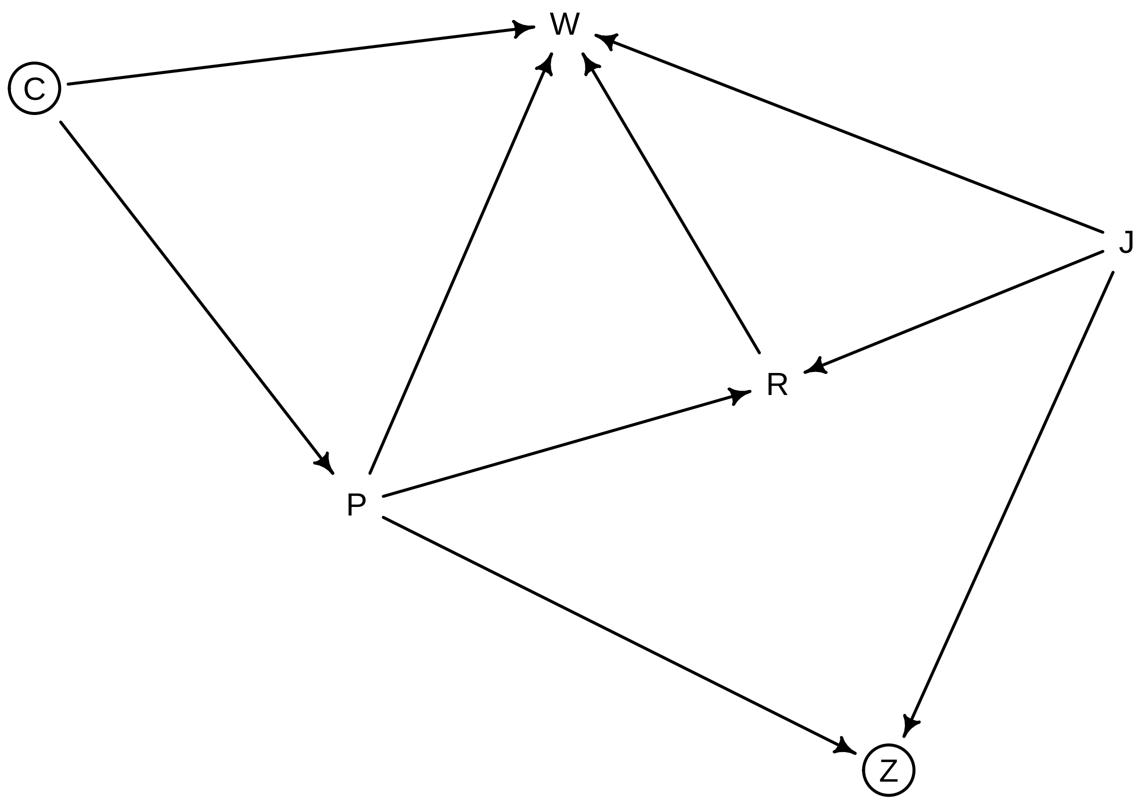
The trankplots above look normal as no chain stays above other chains for a long period of time.

```
##         mean    sd   5.5% 94.5%    n_eff Rhat4
## a     -0.001 0.034 -0.054 0.052 1960.816 0.999
## bR     0.779 0.049  0.701 0.857 1321.351 1.003
## bJ     0.107 0.037  0.047 0.166 1525.805 1.001
## bP    -0.011 0.047 -0.084 0.064 1440.341 1.002
## sigma  0.587 0.024  0.550 0.626 2003.985 1.001
```

```
##         mean    sd   5.5% 94.5%
## a      0.000 0.034 -0.054 0.054
## bR     0.778 0.048  0.701 0.856
## bJ     0.108 0.036  0.050 0.166
## bP    -0.012 0.046 -0.086 0.062
## sigma  0.580 0.024  0.542 0.619
```

The two tables above show a comparison for the mR model, the first table using Markov Chains and the second using quadratic approximation. There is no large variation between the two models and since the other models use the same structure, quadratic approximation is ok to use for the models in this project.

This model is conditioned on walkability scores taken only for locations in the city of Corpus Christi. Other cities or rural areas influence the walkability scores for locations in Corpus, as the original study uses data from the entire US. This model also doesn't account for factors that influence the predictors. One influence that may be of interest is how different areas are used by cities and governments. Often this is referred to as what an area is zoned for and can restrict land to be used for various purposes. For example, a section in the city zoned for residential purposes may differ in population density or job density when compared to another area zoned for industrial or business purposes. A potential DAG could look as follows, where C is city and Z is zoning laws



In this DAG, Z, zoning laws, acts as a collider. If we condition on what that block group is developed for, we may see a relationship between the job and population density. A city, C, can also vary in size which could affect the population density and walkability of certain areas. For example, a downtown in a large, metropolitan city may be easier to walk in compared to a downtown in a rural city.

Conclusion

This project shows that for the city of Corpus Christi, areas with higher concentrations of people, jobs, or road network density indicate that area is more likely to be considered walkable. Population density affects road density and has a larger direct influence on walkability than job density. This information is important to know because people living in walkable areas may benefit from increased exercise resulting in added public health benefits. A further look at influences such as zoning restrictions or including data from other cities or rural areas may be of interest in further studies.

Citations and Data Source

[1], [2] - National Walkability Index Methodology and User Guide

[3] - data.gov

[4] - Smart Location Database Technical Documentation and User Guide

R Code

The following is R code used to create models, plots, and DAGs. Any model can be summarized using 'precis()'

```
# Libraries used
library(tidyverse)
library(rethinking)
```

```r
library(dagitty)

# Create and Draw Directed Acyclic Graph
dag <- dagitty('dag{
               P -> W
               J -> W
               R -> W
               J -> R
               P -> R}')
drawdag(dag)

# Model used to test Job and Population independence
m <- quap(
  alist(
    J ~ dnorm(mu,sigma),
    mu <- a + bp * P,
    a ~ dnorm(0,0.2),
    bp ~ dnorm(0,0.5),
    sigma ~ dexp(1)
  ), data = d)

# Check Adjustment Sets for P as predictor of W
adjustmentSets(dag, outcome = 'W', exposure = 'P')

# Population as predictor of Walkability model
mP <- quap(
  alist(
    W ~ dnorm(mu,sigma),
    mu <- a + bP * P,
    a ~ dnorm(0,0.2),
    bP ~ dnorm(0,0.5),
    sigma ~ dexp(1)
  ), data = d)

# Prior check for model mP
prior <- extract.prior(mP)
mu <- link(mP, post = prior, data = list(P=c(-2,2)))
plot(NULL, xlim = c(-2,2), ylim = c(-2,2),
     xlab = 'Population Density (std)', ylab = 'Walkability (std')
for (i in 1:50) lines(c(-2,2), mu[i,], col = col.alpha('black',0.4))

# Posterior plot for model mP
a.seq <- seq(from = min(d$P), to = max(d$P), length.out = 50)
pop.seq <- list(P = a.seq)
mu <- link(mP, data = pop.seq)
mu.mean <- apply(mu, 2, mean)
mu.HPDI <- apply(mu, 2, HPDI)
sim.walk <- sim(mP, data = pop.seq)
walk.HPDI <- apply(sim.walk, 2, HPDI)
plot(W~P, data = d, col = col.alpha(rangi2,0.5),
     xlab = 'Pop Density (std)', ylab = 'Walkability (std)')
lines(a.seq, mu.mean)
shade(mu.HPDI, a.seq)
```

```r
shade(walk.HPDI, a.seq)

# Check Adjustment Sets for P as predictor of J
adjustmentSets(dag, outcome = 'W', exposure = 'J')

# Create model with job density as predictor of walkability
mJ <- quap(
  alist(
    W ~ dstudent(2,mu,sigma),
    mu <- a + bJ * J,
    a ~ dnorm(0,0.2),
    bJ ~ dnorm(0,0.5),
    sigma ~ dexp(1)
  ), data = d)

# Prior check for model mJ
prior <- extract.prior(mJ)
mu <- link(mJ, post = prior, data = list(J=c(-2,2)))
plot(NULL, xlim = c(-2,2), ylim = c(-2,2),
     xlab = 'Job Density (std)', ylab = 'Walkability (std)')
for (i in 1:50) lines(c(-2,2), mu[i,], col = col.alpha('black',0.4))

# Posterior plot for model mJ
a.seq <- seq(from = min(d$J), to = max(d$J), length.out = 50)
job.seq <- list(J = a.seq)
mu <- link(mJ, data = job.seq)
mu.mean <- apply(mu, 2, mean)
mu.HPDI <- apply(mu, 2, HPDI)
sim.walk <- sim(mJ, data = job.seq)
walk.HPDI <- apply(sim.walk, 2, HPDI)
plot(W~J, data = d, col = col.alpha(rangi2,0.5),
     xlab = 'Job Density (std)', ylab = 'Walkability (std)')
lines(a.seq, mu.mean)
shade(mu.HPDI, a.seq)
shade(walk.HPDI, a.seq)

# Adjustment sets for road density as predictor of walkability
adjustmentSets(dag, outcome = 'W', exposure = 'R')

# Model using P, J, R to predict walkability
mR <- quap(
  alist(
    W ~ dnorm(mu,sigma),
    mu <- a + bR * R + bJ * J + bP * P,
    a ~ dnorm(0,0.2),
    bR ~ dnorm(0,0.5),
    bJ ~ dnorm(0,0.5),
    bP ~ dnorm(0,0.5),
    sigma ~ dexp(1)
  ), data = d)

# Coefficient comparison bewteen models
.pardefault <- par()
```

```r
plot(coeftab(mP,mJ,mR), par=c('bR','bJ','bP'))
par(.pardefault)

# Counterfactual Model Manipulating P
mCF_P <- quap(
  alist(
    # P -> W <- R
    W ~ dnorm(mu, sigma),
    mu <- a + bR*R + bP*P,
    a ~ dnorm(0,0.2),
    bR ~ dnorm(0,0.5),
    bP ~ dnorm(0,0.5),
    sigma ~ dexp(1),
    # P -> R
    R ~ dnorm(muPR, sigmaPR),
    muPR <- aPR + bPR*P,
    aPR ~ dnorm(0,0.2),
    bPR ~ dnorm(0,0.5),
    sigmaPR ~ dexp(1)
  ), data = d)
precis(mCF_P)

# Simulate P values
P.seq <- seq(from=min(d$P), to=max(d$P), length.out = 287)
sim.P <- data.frame(P=P.seq)
sP <- sim(mCF_P, data=sim.P, vars=c('R','W'))

# Plot simulated P
plot(sim.P$P,colMeans(sP$W),ylim=c(min(sP$W),max(sP$W)),type='l',
     xlab='Manipulated P', ylab = 'Counterfactual W')
shade(apply(sP$W,2,PI),sim.P$P)
plot(sim.P$P,colMeans(sP$R),ylim=c(min(sP$R),max(sP$R)),type='l',
     xlab='Manipulated P', ylab = 'Counterfactual R')
shade(apply(sP$R,2,PI),sim.P$P)

# Simulate J values
J.seq <- seq(from=min(d$J), to=max(d$J), length.out = 287)
sim.J <- data.frame(J=J.seq)
sJ <- sim(mCF_J, data=sim.J, vars=c('R','W'))

# Plot simulated J
plot(sim.J$J,colMeans(sJ$W),ylim=c(min(sJ$W),max(sJ$W)),type='l',
     xlab='Manipulated J', ylab = 'Counterfactual W')
shade(apply(sJ$W,2,PI),sim.J$J)
plot(sim.J$J,colMeans(sJ$R),ylim=c(min(sJ$R),max(sJ$R)),type='l',
     xlab='Manipulated J', ylab = 'Counterfactual R')
shade(apply(sJ$R,2,PI),sim.J$J)

# Ulam comparison model
u <- ulam(
  alist(
    W ~ dnorm(mu,sigma),
    mu <- a + bR * R + bJ * J + bP * P,
```

```r
    a ~ dnorm(0,0.2),
    bR ~ dnorm(0,0.5),
    bJ ~ dnorm(0,0.5),
    bP ~ dnorm(0,0.5),
    sigma ~ dexp(1)
  ), data = d, chains = 4, cores = 4)

# Check Ulam model
traceplot(u)
trankplot(u)
precis(u)

# Altered DAG
dag <- dagitty('dag{
               P -> W
               J -> W
               R -> W
               J -> R
               P -> R
               J -> Z <- P
               P <- C -> W
               Z [unobserved]
               C [unobserved]}')
drawdag(dag)
```