Midterm & Final Projects - Time Series

Phoenix Williams

```r
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.6
v forcats   1.0.1     v stringr   1.6.0
v ggplot2   4.0.1     v tibble    3.3.0
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.2.0
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becon
```

```r
library(forecast)
```

```
Registered S3 method overwritten by 'quantmod':
  method            from
  as.zoo.data.frame zoo
```

```r
library(lme4)
```

```
Loading required package: Matrix

Attaching package: 'Matrix'

The following objects are masked from 'package:tidyr':

    expand, pack, unpack
```

```r
library(lmerTest)
```

```
Attaching package: 'lmerTest'

The following object is masked from 'package:lme4':

    lmer

The following object is masked from 'package:stats':

    step
```

```r
library(prophet)
```

```
Loading required package: Rcpp
Loading required package: rlang

Attaching package: 'rlang'

The following objects are masked from 'package:purrr':

    flatten, flatten_chr, flatten_dbl, flatten_int, flatten_lgl,
    flatten_raw, invoke, splice
```

Additional 5027 Questions

Problem 1

With respect to an ARIMA model:

1. Briefly provide a few reasons why maximum likelihood estimation is desirable.

Maximum likelihood estimation is desirable because it is unbiased and fast compared to other estimators.

2. Looking at the equation/formula for how prediction intervals are calculated, briefly explain what in the formula leads to wider prediction intervals as you forecast more time steps into the future.

The uncertainty in a prediction interval for ARIMA forecasting sources primarily from the standard deviation of the forecasting step which is proportional to a sum of the moving average coefficients. Because we sum more moving average coefficients with each forecasting step, the standard deviation of the forecasting step increases with each subsequent step.

Problem 3

Briefly, why is cross-validation important to conduct? What considerations are needed in particular for cross-validation with univariate time series data (that is, a single correlated series of data points)? It is a very non-intuitive concept to be personal to be "forecasting the future". How could cross-validation improve your confidence in being able to accurately predict the future?

Cross validation is important to conduct because it helps establish the generalizability of our model. This is especially important in univariate time series because we are basing our model off of very little data compared to a regression or classification model. Cross validation helps confirm that we have reduced the autocorrelation/partial autocorrelation within out times series data and eliminate the impacts of random anomalies or real world events that may have affected our data to ensure that we are analyzing and forecasting based upon the trend itself rather than other potential factors.

Problem 4

When fitting a machine learning time series model, discuss at least two exact places in the modeling process or decisions an analyst has to make that reflect the trade-off (if any) of efficiency/computational time with accuracy/variance.

Machine learing models involve multiple levels of transformations and running multiple iterations for each parameter in each transformation. Depending upon the size of the data, this can get really computationally expensive very fast, but also additional levels and parameters can increase the accuracy and decrease the variance of our predictions. Therefore, it is important for statisticians to weigh the requirements for accuracy and variance in their work against the limitations on time and computational power and the size of their data before they start working.

Problem 5

Imagine you did this same report for a job and not this class. After reading your report, your boss/manager/client asks you what you would need to make the conclusions stronger.

1. Before reading (2) below, what is your immediate (professional) response to their question? That is, what would you answer or how would you get clarification about what they mean by "what you need" or "make the conclusions stronger"? Just give a few brief ideas so you see how you might react, as of this moment, to this question in real life.

To get clarification on what is meant by "making the conclusions stronger", I would ask for clarity about which metrics are being used to evaluate the strength of the conclusions (i.e.,accuracy, generalizability, precision, etc.) And then I would ask for what resources we can get access to. I would be most interested in getting new or improved data that was verified with a clear primary source, given in numbers rather than percentages, and given on a monthly basis rather than an annual basis (to perhaps study any patterns of seasonality in the phenomenon), increased specificity for age groups and other idenities, and for a longer stretch of time. Depending upon how large and complex the new data was, I may then be in need of additional computational power. I would also perhaps ask for support staff to scan the full spectrum of methods available to use. There would also be a not irrelevant question of time.

2. Now, for (2(1)) and (2(2)) below, what is your answer if the boss wants to provide you any needed resources (computation, support staff, new data, … anything you want to ask for) to:

    1. Increase the accuracy of your conclusion

    To increase the accurcacy of my conclusion, I would be most interested in better data, particularly in terms of data for a longer period of time that is given on a more frequent interval and data given in numbers rather than percentage values.

    2. increase the generalizability of your conclusion.

    To increase the generalizability of my conclusions, I would be interested in doing what we can to get more specific data, whether that is through the period of the data, the demographic association of those in the dataset, or the verification of the source of the data.

A Time Series Analysis of Transgender Suicide Rates

Suicide is a leading cause of death worldwide, yet its risk is not distributed evenly across populations. Transgender and gender-diverse individuals experience disproportionately high rates of suicidal ideation and attempts, which has drawn increasing attention from researchers, clinicians, policymakers, and the LGBTQ+ community. These disparities are not inherent to being transgender, but instead reflect the cumulative effects of social stigma, discrimination, barriers to affirming healthcare, and exposure to violence or rejection. Understanding suicide risk among transgender populations is therefore not only a matter of mental health research, but also a broader question about social environments, policy decisions, and access to support.

Examining these patterns carefully and responsibly is essential for developing interventions that reduce harm and promote well-being.

From a statistical perspective, understanding whether suicide-related outcomes among transgender individuals are continuing to rise, stabilizing, or accelerating over time requires methods capable of capturing long-term trends while remaining sensitive to potential structural changes. In addition to classical time-series approaches, modern trend-focused models such as Prophet offer a complementary framework for this task by explicitly modeling gradual trend evolution and possible changes in slope over time. Including these models alongside more traditional approaches allows for a richer assessment of whether observed increases reflect steady long-term growth or more recent accelerations that may correspond to shifts in social or policy contexts.

Research Question

The central policy question motivating this analysis is whether suicide ideation and attempt rates among transgender individuals in the United States have continued to increase over time, and whether recent changes represent a continuation of long-term trends or meaningful deviations that warrant urgent intervention. Answering this question has direct real-world implications for public health planning, mental health treatment, and the evaluation of policies related to healthcare access, discrimination, and social support for transgender populations.

Using annual time-series data from 2000 to 2022, this study models long-term trends in both suicidal ideation and suicide attempts, assesses the rate at which these outcomes have changed over time, and examine whether periods of accelerated increase or structural shifts are present. A combination of classical time-series models (including smoothing methods and ARIMA) and modern trend-based approaches (specifically, two variants of Prophet) is employed to evaluate whether conclusions about temporal change are robust across modeling frameworks. Prophet models are particularly valuable in this setting because they allow explicit exploration of trend flexibility and potential changepoints, complementing the stricter assumptions required by ARIMA models.

Data

The data used in this analysis are appropriate for addressing the research question because they directly measure the outcomes of interest—rates (percentages) of suicidal ideation and suicide attempts among transgender individuals in the United States—over an extended period of time. Since the goal is to understand whether these rates exhibit long-term trends or meaningful changes over time, a longitudinal dataset spanning multiple decades is necessary. The data provides annual estimates from 2000 through 2022, allowing for the examination of gradual changes, potential accelerations, and broader temporal patterns.

A key limitation of the data is that they are based on aggregatd survey estimates rather than administrative records, which means they may be affected by reporting bias, changes in survey methodology, or evolving social willingness to disclose suicidal thoughts or attempts. However, these limitations also motivate the use of multiple modeling approaches. In particular, Prophet models are well suited for such data because they emphasize trend estimation and changepoint detection rather than short-term dependence, making them appropriate for relatively short, annual time series with potential structural shifts.

The data were obtained from **Statista**, which aggregates and standardizes data from nationally representative surveys and reputable research sources. According to Statista's description, the estimates reflect the percentage of transgender individuals in the United States who have considered suicide and the percentage who have attempted suicide in each year. Statista serves as the secondary data source, and its compilation methods allow for consistent comparison across time, though this also means the analysis is dependent on the quality and consistency of the original surveys.

The resulting time series consists of 23 annual observations (2000–2022) for each outcome variable. The time step is one year, and the primary variables of interest are the percentage of transgender individuals who have considered suicide and the percentage who have attempted suicide. The series is relatively short by time-series standards, which constrains the complexity of models that can be reliably estimated. This characteristic informed the modeling strategy, favoring interpretable trend-based methods (such as Holt's linear trend and Prophet) and cautious use of more assumption-heavy models like ARIMA.

Because the dataset was obtained from a published summary rather than a downloadable file, the values were manually entered into R and structured as an annual time series with one observation per year from 2000 to 2022. The data were then converted into time-series objects to facilitate trend modeling and other longitudinal analyses.

```r
# Recreate transgender suicide rate data (U.S., 2000-2022)
trans_suicide <- data.frame(
  year = 2000:2022,
  considered_pct = c(
    61, 63, 66, 65, 64, 66, 69, 68, 70, 73, 79, 74, 77,
    75, 79, 78, 76, 77, 78, 79, 83, 79, 80
  ),
  attempted_pct = c(
    28, 29, 31, 32, 31, 32, 33, 34, 35, 35, 41, 37, 38,
    40, 39, 37, 38, 39, 40, 39, 44, 39, 40
  )
)

trans_suicide <- trans_suicide[order(trans_suicide$year),]
# Convert to time-series objects
```

```r
considered_ts <- ts(trans_suicide$considered_pct, start = 2000, frequency = 1)
attempted_ts  <- ts(trans_suicide$attempted_pct,  start = 2000, frequency = 1)

# Quick check
head(trans_suicide)
```

```
  year considered_pct attempted_pct
1 2000             61            28
2 2001             63            29
3 2002             66            31
4 2003             65            32
5 2004             64            31
6 2005             66            32
```

Statistical Methods

The analysis will proceed in several structured stages designed to move from understanding the data to drawing robust conclusions about long-term trends in transgender suicide ideation and attempt rates. First, exploratory data analysis (EDA) will be conducted to summarize key features of each time series. This will include descriptive statistics, inspection of overall trends, and visualization of the series over time. These exploratory steps are essential for identifying dominant patterns (e.g., monotonic increases, possible inflection points, or periods of accelerated change) and for motivating subsequent modeling choices, including the appropriateness of trend-based or autoregressive models are appropriate.

In light of the models retained for analysis, additional exploratory emphasis will be placed on **trend shape and stability over time**. In particular, the EDA will include careful visual inspection for potential changes in slope that could motivate more flexible trend models. This expanded exploratory focus directly supports the use of Holt's linear trend models and Prophet, both of which allow for evolving trends rather than assuming a constant rate of change across the entire study period.

Following exploratory data analysis, the first modeling stage will focus on non-ARIMA models that emphasize long-term trend characterization. First, a **linear mixed-effects model** will be fit to the combined dataset containing both outcomes (considered and attempted suicide rates), with year treated as a fixed effect and outcome type included to account for systematic differences in level between ideation and attempts. This model leverages the shared temporal structure of the two related outcomes which allowing for meaningful differences between them. In addition, **Holt's linear trend model** will be estimated for each series both **with and without damping.** These smoothing-based models are well suited for short, annual time series with strong trends and no seasonality, and they provide interpretable estimates of how the level and slope of the series evolve over time. Comparing damped and undamped versions allows assessment of whether trends appear to be continuing at a steady rate or beginning to flatten.

Next, **ARIMA models** will be estimated for each series to formally assess temporal dependence and short-term dynamics. Prior to ARIMA estimation, assumptions such as approximate stationarity and stable variance will be evaluated and addressed as needed through differencing. Autocorrelation and partial autocorrelation functions will be examined to guide the selection of autoregressive and moving average orders, and candidate models will be compared using information criteria and residual diagnostics. Residual analysis will be used to validate the final ARIMA specifications and ensure that remaining temporal structure has been adequately captured.

Finally, **Prophet models** will be incorporated as flexible, trend-focused alternatives that complement classical smoothing and ARIMA approaches. Two Prophet specifications will be estimated for each outcome: a baseline model with default trend smoothness and a second variant with a more flexible changepoint prior. These models will be used to assess whether the data support a smoothly evolving long-term trend or exhibit evidence of structural changes in slope that may corresponse to broader social or policy shifts. Prophet's explicit trend decomposition and changepoint framework make it particularly valuable for addressing the core research question of whether increases in suicide-related outcomes represent steady long term growth or periods of acceleration.

To evaluate the robustness of model-based conclusions and guard against overfitting, at least one model in this analysis will be assessed using a formal cross-validation procedure. Given the short length and annual resolution of the time series, traditional random cross-validation is not appropriate, as it would break the temporal ordering of the data. Instead, **time-series cross-validation** will be employed, in which models are trained on an initial contiguous block of observations and evaluated on subsequent observations. This approach preserves the chronological structure of the data and more closely reflects the real-world task of using past information to explain or forecast future outcomes.

Time-series cross-validation will be applied primarily to the **Prophet models**, which are well suited to rolling-origin evaluation due to their flexible trend structure and forecasting framework. Specifically, the models will be refit repeatedly on expanding training windows and evaluated on held-out future observations, with forecast accuracy summarized using appropriate error metrics. Comparing cross-validated performance across the baseline and flexible-changepoint Prophet specifications provides insight into whether increased trend flexibility improves out-of-sample performance or leads to overfitting. Because ARIMA and Holt-based models are more sensitive to small sample sizes and parameter uncertainty in short annual series, cross-validation results are interpreted cautiously and used primarily as a comparative diagnostic rather than a definitive model selection criterion. Overall, incorporating time-series cross validation strengthens the analytical framework by ensuring that key modeling conclusions are supported not only by in-sample fit, but also by performance on unseen data.

Together, these steps form a coherent and comprehensive analytical framework. Exploratory data analysis establishes the empirical context and motivates model choice; non-ARIMA models such as linear mixed effect ans Holt's methods capture broad, interpretable trends; ARIMA models formally evaluate autocorrelation and short-term dynamics; and Prophet models provide flexible assessments of trend evolution and potential structural change. This layered approach ensures that conclusions about long-term increases and temporal structure are not driven by a single modeling assumption, but instead are supported by multiple perspectives that are both statistically appropriate for the data and substantively relevant to the research question.

Exploratory Data Analysis

Minimal data cleaning was required because the dataset was already structured as a time series with one observation per year. Preparation steps primarily involved verifying the chronological ordering of the data, ensuring consistent units across years, and separating the two outcome variables—suicidal ideation and suicide attempts—so they could be analyzed both jointly and independently. No interpolation or imputation was performed, as the dataset contained complete annual observations for the study period. A conceptual decision was made to treat the series as annual rather than attempting finer temporal resolution, since the underlying data do not support sub-annual measurement.

Summary statistics and descriptive features

As part of the exploratory analysis, summary statistics such as the minimum, maximum, mean, and range of each time series were computed. These summaries help contextualize the magnitude of observed changes over time and highlight the substantial increase in both suicidal ideation and suicide attempts across the study period. The relatively short length of the series (23 annual observations) is noted as a constraint that favors parsimonious models and careful interpretation, particularly for models that require estimating multiple parameters.

```
# Summary statistics for suicidal ideation
summary(considered_ts)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   61.0    67.0    75.0    73.0    78.5    83.0 
```

```
# Summary statistics for suicide attempts
summary(attempted_ts)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  28.00   32.50   37.00   36.13   39.00   44.00 
```

```
# Range calculations
range(considered_ts)
```

```
[1] 61 83
```

```
range(attempted_ts)
```

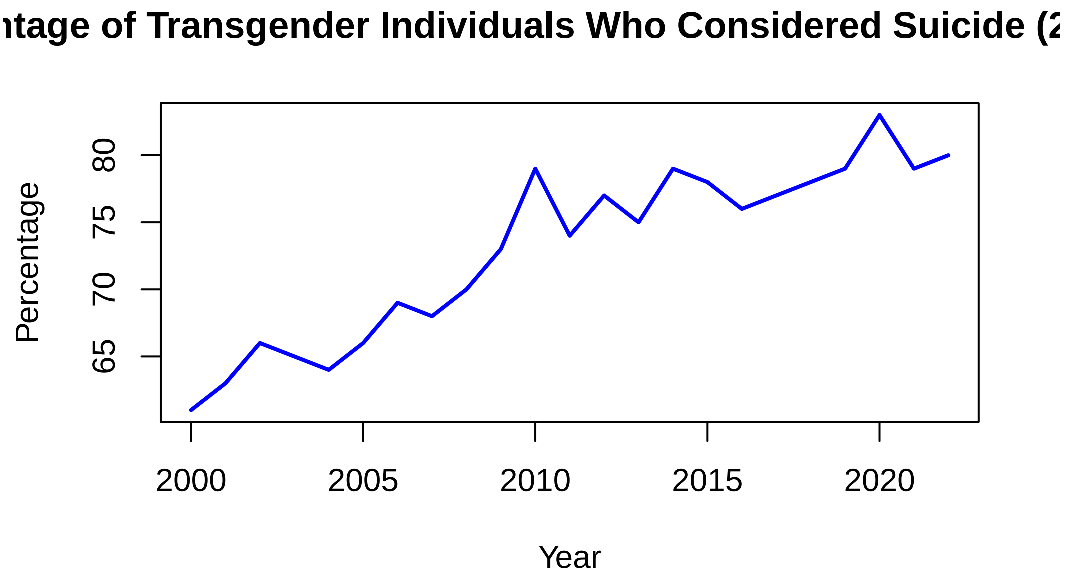
```
[1] 28 44
```

These statistics confirm that both outcomes exhibit large absolute changes over time rather than small fluctuations around a stable mean, reinforcing the need for models that explicitly account for trend.

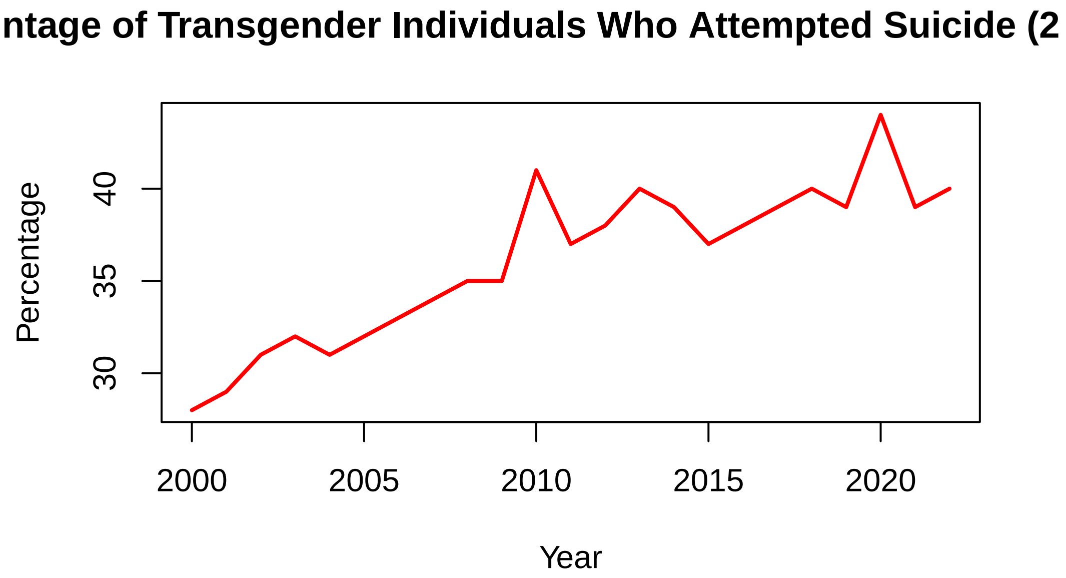
Exploratory visualization

Time-series plot were produced for each outcome variable, displaying the percentage of transgender individuals who have considered suicide and attempted suicide from 2000 to 2022. These visualizations show a clear upward trend in both series, with no obvious seasonal structure due to the annual time step. While there are year to year fluctuations and occasional spikes, there is no evidence of repeating cyclical patterns. The absence of seasonality supports the use of non-seasonal models, while the strong upward trajectory indicates that any autoregressive modeling must address nonstationarity, likely through detrending or differencing.

```
# Plot suicidal ideation over time
plot(
  considered_ts,
  main = "Percentage of Transgender Individuals Who Considered Suicide (2000-2022)",
  ylab = "Percentage",
  xlab = "Year",
  col = "blue",
  lwd = 2
)
```



```
# Plot suicide attempts over time
plot(
  attempted_ts,
  main = "Percentage of Transgender Individuals Who Attempted Suicide (2000-2022)",
  ylab = "Percentage",
  xlab = "Year",
  col = "red",
  lwd = 2
)
```



Visual inspection of these plots guided subsequent modeling decisions by confirming that trend-based models, such as Holt's linear trend models and Prophet, are appropriate initial tools and that ARIMA modeling will require explicit treatment of nonstationarity.

Preparation of the Time Series: Components and Anomalies

Before fitting time-series models, the two outcome variables-percentage of transgender individuals who have considered suicide and percentage who have attempted suicide-were examined for key structural components including trend and seasonality as well as for potential anomalies. Because the data are annual, covering the period from 2000–2022, no within-year seasonal pattern was expected. This expectation was confirmed empirically, as both series have a frequency of one and show no evidence of recurring cyclical behavior.

Visual inspection of the time-series plots confirmed that both outcomes are dominated by a strong long-term upward trend in both outcomes, with year-to-year fluctuations that are small relative to the overall magnitude of change. Given that the data are aggregated annually and derived from population-level survey estimates, no observations were removed or adjusted as anomalies, Instead, any apparent short-term deviations were treated as meaningful reflections of real-world variability rather than measurement error. This assessment supports the use of trend-focused models, such as Holt's linear trend models and Prophet, and informed later decisions regarding detrending and differencing for ARIMA modeling.

```
# Check for seasonality (should be none for annual data)
frequency(considered_ts)
```

```
[1] 1
```

```
frequency(attempted_ts)
```

```
[1] 1
```

```
# Outlier/anomaly check (treat as diagnostic only; do NOT automatically remove points)
out_considered <- try(tsoutliers(considered_ts), silent = TRUE)
out_attempted  <- try(tsoutliers(attempted_ts), silent = TRUE)

out_considered
```

```
$index
integer(0)

$replacements
numeric(0)
```

```
out_attempted
```

```
$index
[1] 11 21

$replacements
[1] 36 39
```

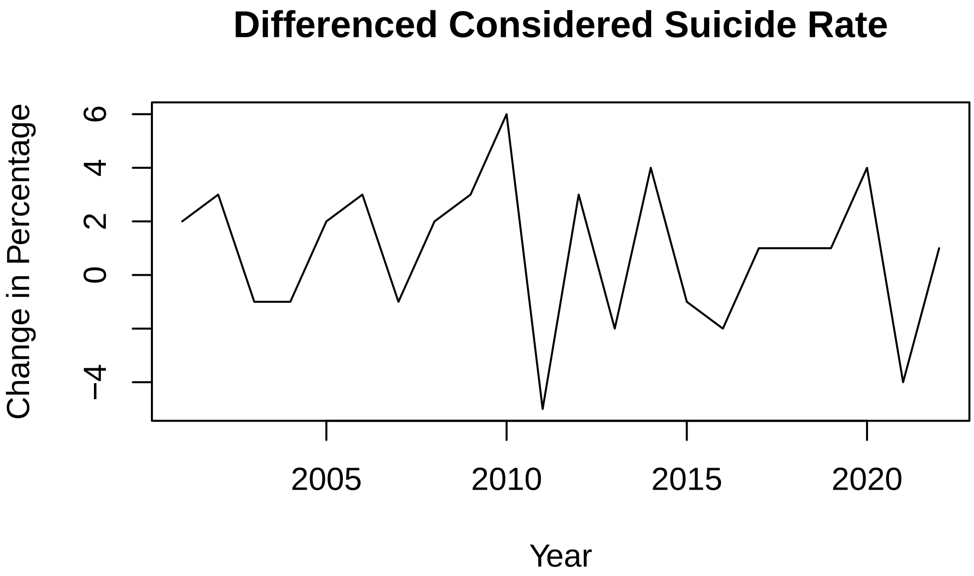
Preliminary autocorrelation assessment

Because ARIMA models rely on temporal dependence, exploratory autocorrelation (ACF) and partial autocorrelation (PACF) analyses were conducted to assess whether meaningful serial correlation exists beyond the dominant trend. Initial inspection of the original series indicated strong autocorrelation at low lags, a pattern consistent with nonstationary data driven by long-term trends rather than stable short-term dependence.

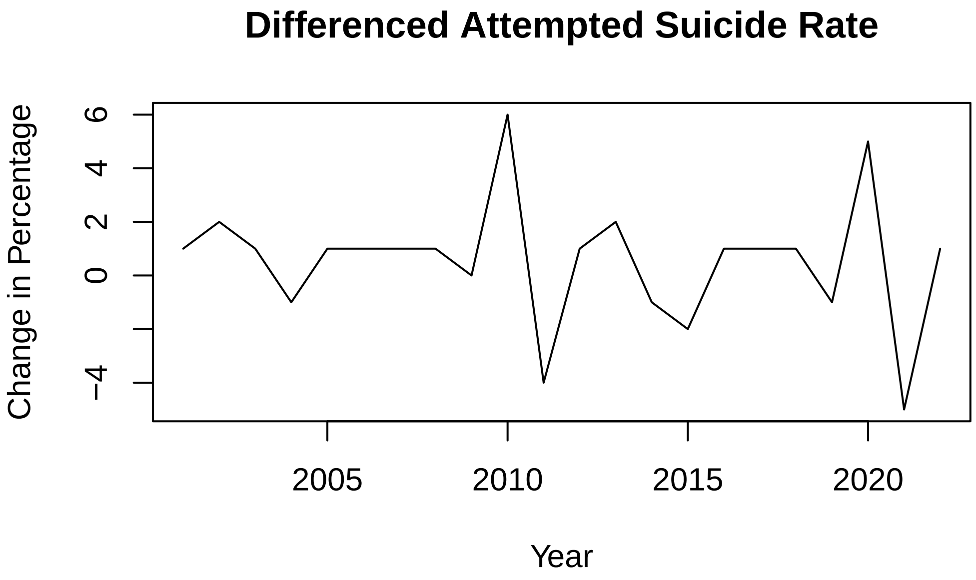
Prior to fitting ARIMA models, the series were evaluated for stationarity, a fundamental assumption of ARIMA modeling. Visual inspection and summary statistics suggested pronounced nonstationarity in both outcomes due to their sustained upward trajectories. As a result, first differencing was applied to remove the trend and stabilize the mean. The differenced series were then examined to assess whether approximate stationarity had been achieved and whether additional transformations were necessary. Because the data are annual, no seasonal differencing or SARIMA terms were considered.

Autocorrelation (ACF) and partial autocorrelation (PACF) plots of the differenced series were generated to evaluate the remaining temporal structure and to guide the selection of candidate autoregressive and moving average terms in subsequent ARIMA models.

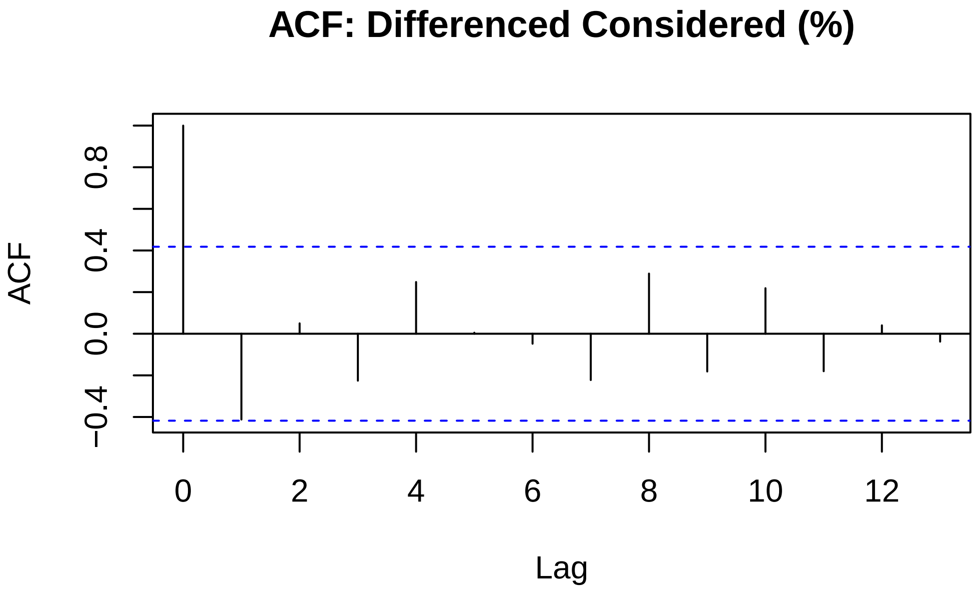
```
# Check stationarity visually and with a simple differencing step
d_considered <- diff(considered_ts, differences = 1)
d_attempted  <- diff(attempted_ts,  differences = 1)

plot(d_considered, main = "Differenced Considered Suicide Rate", ylab = "Change in Percentag
```



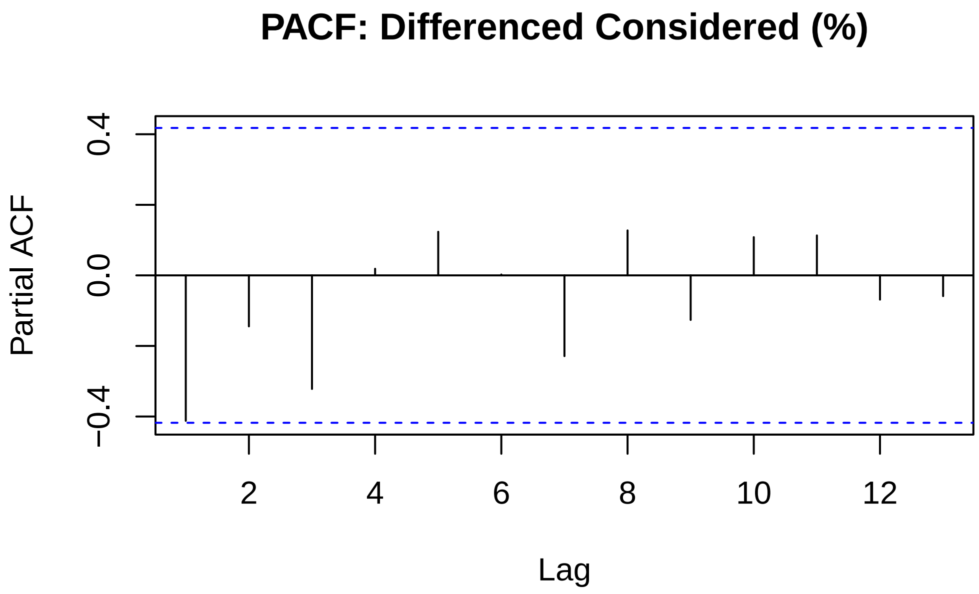
```
plot(d_attempted,  main = "Differenced Attempted Suicide Rate",  ylab = "Change in Percentag
```



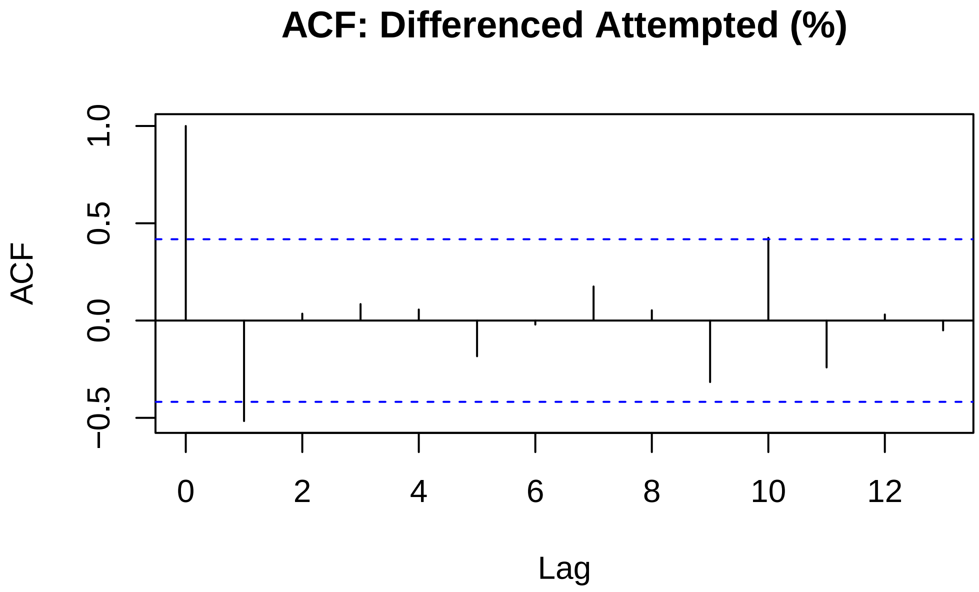
```r
# ACF/PACF for candidate p and q selection
acf(d_considered, main = "ACF: Differenced Considered (%)")
```



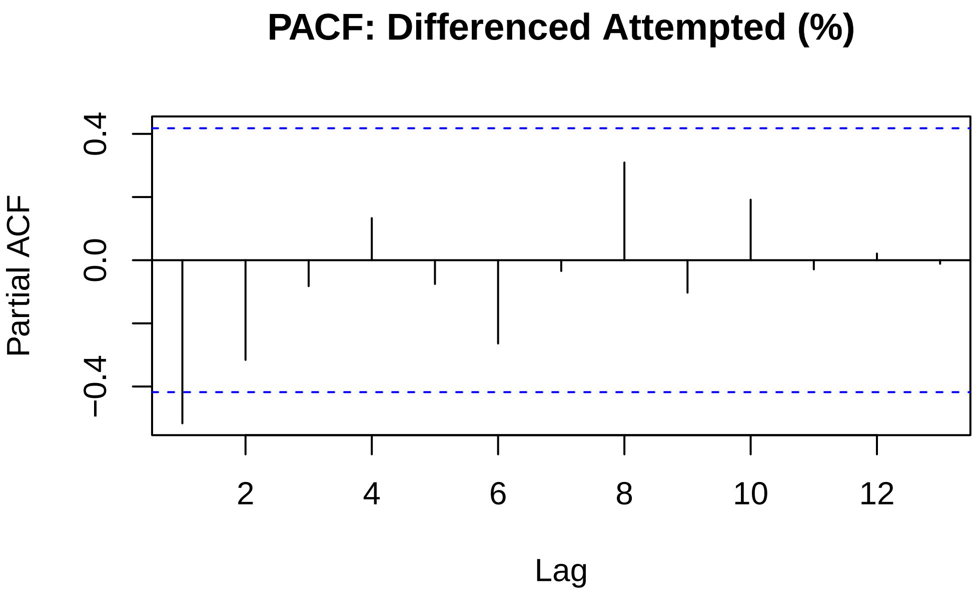
```r
pacf(d_considered, main = "PACF: Differenced Considered (%)")
```



```r
acf(d_attempted, main = "ACF: Differenced Attempted (%)")
```



```
pacf(d_attempted, main = "PACF: Differenced Attempted (%)")
```



For both datasets, we see that the ACF significance lag is at least one and no more than 2 while the PACF significance lag is no more than one.

Together, these exploratory diagnostics establish that the original series are nonstationary and trend-dominated, justifying the use of differencing prior to ARIMA modeling and reinforcing the suitability of trend bacsed approaches such as Holt's linear trend models and Prophet. This preparatory analysis ensures that subsequent modeling decisions are grounded in the observed structure of the data rather than imposed assumptions.

Statistical Analysis

Linear Mixed-Effects Model

A linear mixed-effects model was first fit to the combined dataset containing both outcomes (rates of suicidal ideation and suicide attempts) to provide a joint, interpretable summary of long-term trends. Year was included as a fixed effect to estimate overall temporal change, and outcome type was included as a fixed effect to account for systematic differences in level between ideation and attempts. A random intercept for outcome type to allowed each series to have its own baseline rate while sharing a common temporal structure.

This model serves as a parsimonious baseline that leverages information across both outcomes without imposing identical trajectories. Its primary value lies in quantifying the overall direction and magnitude of change over time while explicitly acknowledging that ideation and attemps are related but distinct phenomena.

```r
# Create long (stacked) dataset: one row per year per outcome
trans_long <- trans_suicide |>
  pivot_longer(cols = c(considered_pct, attempted_pct),
               names_to = "outcome",
               values_to = "pct") |>
  mutate(
    outcome = factor(outcome,
                     levels = c("attempted_pct", "considered_pct"),
                     labels = c("attempted", "considered")),
    year_c = year - 2000
  )

# Mixed model: pct ~ year + outcome + (1|outcome)
# Random intercept by outcome allows different baseline levels for attempted vs considered.
lmm <- lmer(pct ~ year_c + outcome + (1 | outcome), data = trans_long)
```

```
Warning in as_lmerModLT(model, devfun): Model may not have converged with 1
eigenvalue close to zero: 8.8e-10
```

```r
summary(lmm)
```

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: pct ~ year_c + outcome + (1 | outcome)
   Data: trans_long

REML criterion at convergence: 212.3

Scaled residuals:
     Min       1Q   Median       3Q      Max 
-1.80400 -0.58642  0.01423  0.46076  2.76781 

Random effects:
 Groups   Name        Variance Std.Dev.
 outcome  (Intercept) 69.984   8.366   
 Residual              5.904   2.430   
```

```
Number of obs: 46, groups:  outcome, 2

Fixed effects:
                  Estimate Std. Error       df t value Pr(>|t|)    
(Intercept)       28.15217    8.40203 43.00000   3.351  0.00169 ** 
year_c             0.72530    0.05401 43.00000  13.429  < 2e-16 ***
outcomeconsidered 36.86957   11.85253 43.00000   3.111  0.00331 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) year_c
year_c      -0.071       
outcmcnsdrd -0.705  0.000
```

Holt's Linear Trend Models

Holt's linear trend models were estimated separately for each outcome to capture evolving level and trend components without imposing a fixed functional form. These models are well suited to short, annual series with strong trends and no seasonality, making them appropriate for the data at hand.

The undamped specification assumes that recent trend estimates continue indefinitely, while the damped specification allows the trend to gradually flatten. Comparing these two versions provides insight into whether the observed increases appear sustained or show evidence of slowing in recent years-an interpretation that is directly relevant in a public health and policy context.

```r
holt_considered <- holt(considered_ts, damped = FALSE, h = 3)
holt_attempted  <- holt(attempted_ts,  damped = FALSE, h = 3)

holt_considered
```

```
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2023       81.92853 78.72182 85.13524 77.02429 86.83277
2024       82.79344 79.34365 86.24322 77.51745 88.06943
2025       83.65835 79.98141 87.33528 78.03496 89.28174
```

```r
holt_attempted
```

```
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2023       40.21801 37.56550 42.87053 36.16134 44.27468
2024       40.17023 37.47790 42.86256 36.05267 44.28780
2025       40.12245 37.34261 42.90228 35.87106 44.37384
```

```
holt_damped_considered <- holt(considered_ts, damped = TRUE, h = 3)
holt_damped_attempted  <- holt(attempted_ts,  damped = TRUE, h = 3)

holt_damped_considered
```

```
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2023       81.61363 78.70258 84.52468 77.16157 86.06569
2024       82.07688 79.16583 84.98792 77.62481 86.52894
2025       82.51911 79.60806 85.43015 78.06705 86.97117
```

```
holt_damped_attempted
```

```
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2023       41.02955 38.75959 43.29951 37.55795 44.50116
2024       41.23654 38.96657 43.50650 37.76493 44.70814
2025       41.42969 39.15973 43.69966 37.95809 44.90130
```

ARIMA Modeling and Diagnostics

ARIMA Model Estimation and Diagnostics

ARIMA models were estimated for each outcome to assess whether short-term temporal dependence remains after accounting for long-term trends. Based on the exploratory analysis, both series were first differenced to address nonstationarity. Candidate ARIMA specifications were selected using a combination of ACF/PACF diagnostics and information criteria, with automated selection via AICc for a baseline and manual selection based upon our EDA.

Model adequacy was evaluated through residual diagnostics, including checks for remaining autocorrelation and approximate white-noise behavior. These diagnostics are essential to ensure that the ARIMA models have adequately captured the temporal dependence structure and that remaining residuals do not exhibit systematic patterns.

```
# Automated ARIMA selection (AICc)
arima_considered_auto <- auto.arima(considered_ts, seasonal = FALSE, stepwise = FALSE, appro
arima_attempted_auto  <- auto.arima(attempted_ts,  seasonal = FALSE, stepwise = FALSE, appro

summary(arima_considered_auto)
```

```
Series: considered_ts
ARIMA(0,1,1) with drift

Coefficients:
          ma1   drift
      -0.5712  0.8648
s.e.   0.1932  0.2272

sigma^2 = 5.955:  log likelihood = -49.99
AIC=105.98   AICc=107.32   BIC=109.26

Training set error measures:
                     ME     RMSE      MAE        MPE     MAPE      MASE
Training set 0.06168989 2.275489 1.786314 0.08323878 2.394666 0.7414889
                   ACF1
Training set 0.00846492
```

```r
summary(arima_attempted_auto)
```

```
Series: attempted_ts
ARIMA(0,1,1) with drift

Coefficients:
          ma1   drift
      -0.6647  0.5557
s.e.   0.1785  0.1499

sigma^2 = 3.962:  log likelihood = -45.6
AIC=97.21   AICc=98.54   BIC=100.48

Training set error measures:
                     ME     RMSE      MAE       MPE     MAPE      MASE
Training set 0.07835931 1.856175 1.299618 0.2088885 3.425632 0.7147901
                   ACF1
Training set -0.0798807
```

```r
# Example manual ARIMA fits (edit p,d,q based on your ACF/PACF)
# Common starting point for trending annual series is d=1
arima_considered_manual <- Arima(considered_ts, order = c(2,1,1), seasonal = c(0,0,0))
arima_attempted_manual  <- Arima(attempted_ts,  order = c(2,1,1), seasonal = c(0,0,0))

summary(arima_considered_manual)
```

```
Series: considered_ts
ARIMA(2,1,1)

Coefficients:
          ar1      ar2     ma1
      -1.1296  -0.1305  0.9818
s.e.   0.2328   0.2322  0.1283

sigma^2 = 7.547:  log likelihood = -52.4
AIC=112.81   AICc=115.16   BIC=117.17

Training set error measures:
                    ME     RMSE      MAE      MPE     MAPE      MASE       ACF1
Training set 0.9174097 2.496838 2.027984 1.245711 2.747698 0.8418046 -0.1956257
```

```r
summary(arima_attempted_manual)
```

```
Series: attempted_ts
ARIMA(2,1,1)

Coefficients:
          ar1      ar2     ma1
      -0.6961  -0.2467  0.1961
s.e.   0.6771   0.3561  0.6703

sigma^2 = 5.25:  log likelihood = -47.98
AIC=103.95   AICc=106.31   BIC=108.32

Training set error measures:
                    ME     RMSE      MAE      MPE     MAPE      MASE       ACF1
Training set 0.8458201 2.082441 1.507352 2.296925 4.044089 0.8290437 -0.2106624
```

```r
# Coefficient tables include standard errors; interpret p-values where available
coeftest_considered_auto <- try(lmtest::coeftest(arima_considered_auto), silent = TRUE)
coeftest_attempted_auto  <- try(lmtest::coeftest(arima_attempted_auto),  silent = TRUE)
coeftest_considered_manual <- try(lmtest::coeftest(arima_considered_manual), silent = TRUE)
coeftest_attempted_manual  <- try(lmtest::coeftest(arima_attempted_manual),  silent = TRUE)
coeftest_considered_auto
```

```
z test of coefficients:
```

```
      Estimate Std. Error z value  Pr(>|z|)
ma1   -0.57124    0.19323 -2.9563 0.0031136 **
drift  0.86478    0.22715  3.8070 0.0001406 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
coeftest_attempted_auto
```

```


z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
ma1   -0.66474    0.17853 -3.7235 0.0001965 ***
drift  0.55566    0.14990  3.7069 0.0002098 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

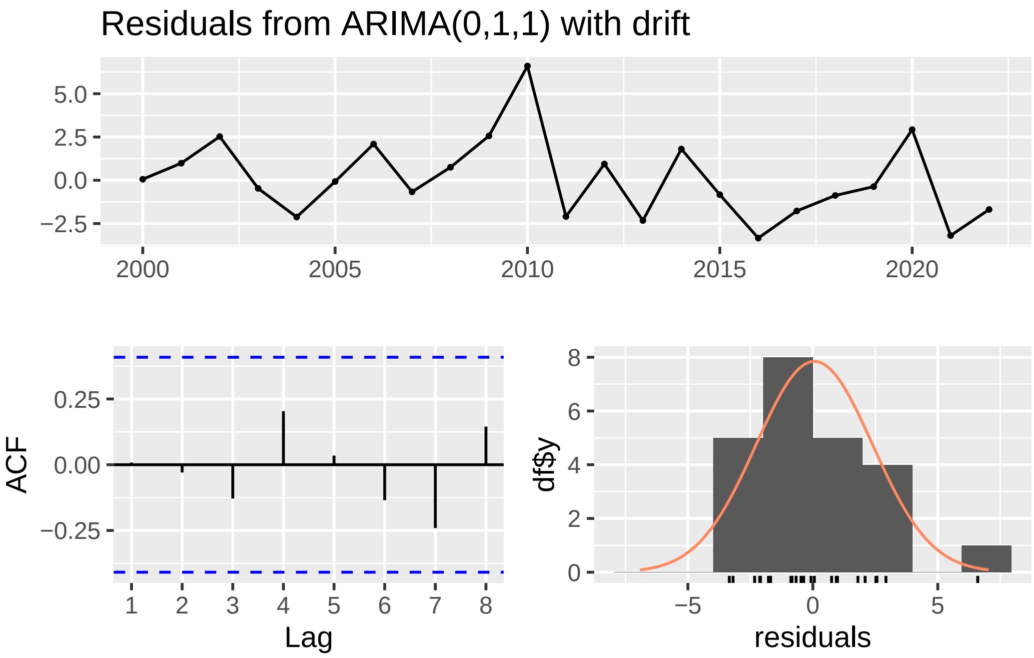
```
coeftest_considered_manual
```

```


z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)
ar1 -1.12956    0.23278 -4.8524 1.219e-06 ***
ar2 -0.13053    0.23220 -0.5622     0.574
ma1  0.98177    0.12827  7.6539 1.949e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
coeftest_attempted_manual
```

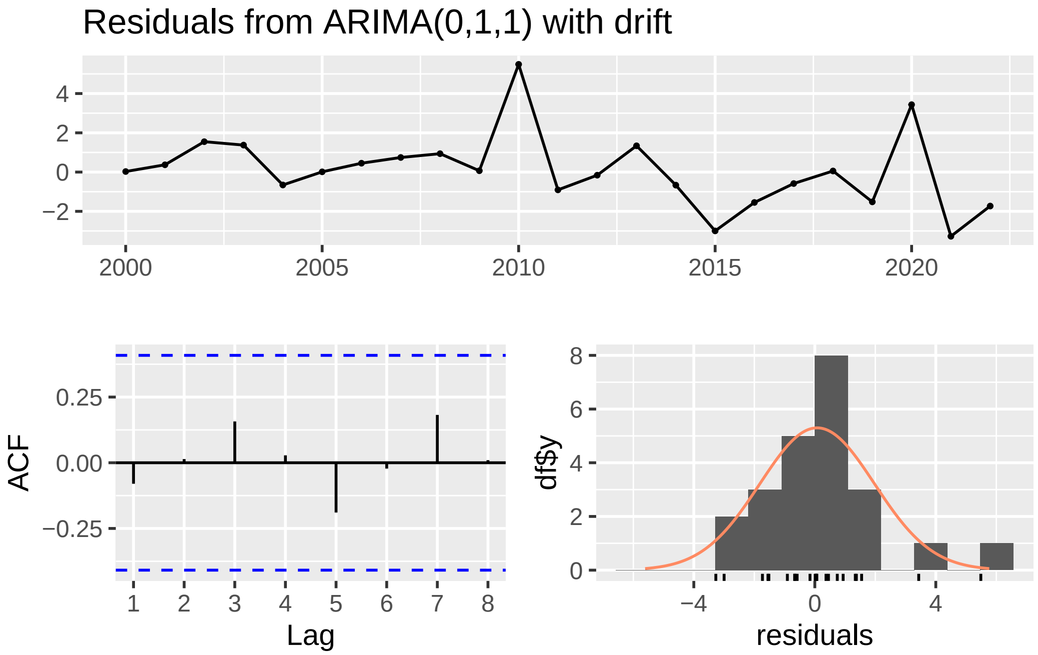
```


z test of coefficients:

    Estimate Std. Error z value Pr(>|z|)
ar1 -0.69612    0.67712 -1.0281   0.3039
ar2 -0.24671    0.35610 -0.6928   0.4884
ma1  0.19608    0.67032  0.2925   0.7699
```

```
# Residual diagnostics
checkresiduals(arima_considered_auto)
```



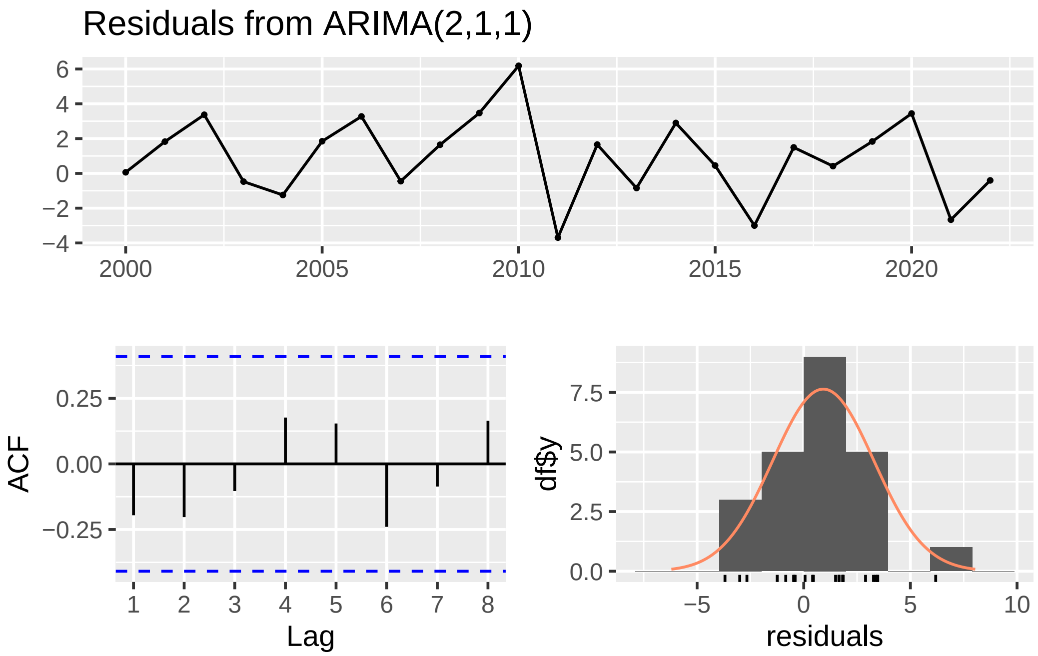
```

	Ljung-Box test

data:  Residuals from ARIMA(0,1,1) with drift
Q* = 1.7975, df = 4, p-value = 0.7729

Model df: 1.   Total lags used: 5
```

```
checkresiduals(arima_attempted_auto)
```



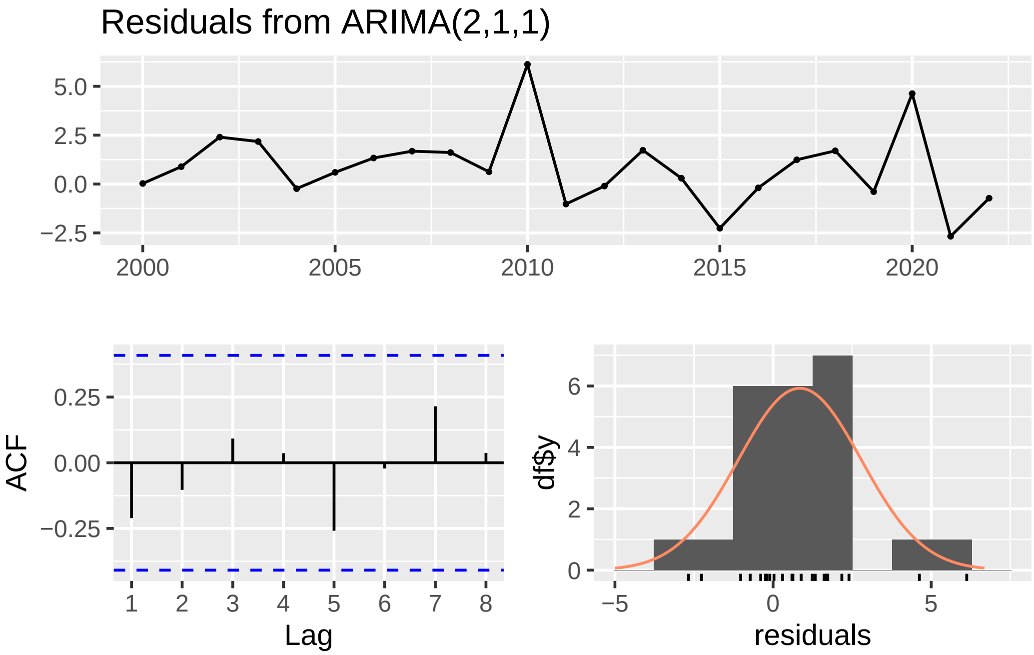
```
    Ljung-Box test

data:  Residuals from ARIMA(0,1,1) with drift
Q* = 2.0532, df = 4, p-value = 0.726

Model df: 1.   Total lags used: 5
```

```
checkresiduals(arima_considered_manual)
```



```
	Ljung-Box test

data:  Residuals from ARIMA(2,1,1)
Q* = 6.076, df = 3, p-value = 0.108

Model df: 3.   Total lags used: 6
```

```r
checkresiduals(arima_attempted_manual)
```



```
	Ljung-Box test

data:  Residuals from ARIMA(2,1,1)
Q* = 3.8885, df = 3, p-value = 0.2738

Model df: 3.   Total lags used: 6
```

```
# Ljung-Box test explicitly (should be non-significant if residuals are white noise)
Box.test(residuals(arima_considered_auto), lag = 10, type = "Ljung-Box")
```

```
	Box-Ljung test

data:  residuals(arima_considered_auto)
X-squared = 6.0558, df = 10, p-value = 0.8106
```

```
Box.test(residuals(arima_attempted_auto),  lag = 10, type = "Ljung-Box")
```

```
	Box-Ljung test
```

```
data:  residuals(arima_attempted_auto)
X-squared = 8.7479, df = 10, p-value = 0.5562
```

```
Box.test(residuals(arima_considered_manual), lag = 10, type = "Ljung-Box")
```

```

    Box-Ljung test

data:  residuals(arima_considered_manual)
X-squared = 7.933, df = 10, p-value = 0.6354
```

```
Box.test(residuals(arima_attempted_manual),  lag = 10, type = "Ljung-Box")
```

```

    Box-Ljung test

data:  residuals(arima_attempted_manual)
X-squared = 11.754, df = 10, p-value = 0.3018
```

While ARIMA models provide a formal framework for modeling serial dependence, their reliance on differencing can obscure long-term trend interpretation, particularly when the primary scientific question concerns sustained increases over decades rather than short-term fluctuations.

Prophet Model and Cross Validation

Insights from the exploratory analysis and ARIMA modeling motivated the inclusion of Prophet as a complementary approach. Visual inspection of the series revealed strong, persistent upward trends with possible changes in slope over time—features that are not easily captured by ARIMA models after differencing. Prophet's explicit trend decomposition and changepoint framework make it well suited for modeling such behavior without requiring stationarity.

Two Prophet specifications were estimated for each outcome: a baseline model with default trend smoothness and a second variant with an increased changepoint prior scale, allowing greater flexibility in trend evolution. Comparing these specifications provides insight into whether the data support smooth long-term growth or exhibit evidence of accelerated change.

```r
# Prepare data for Prophet with fixed changepoints
prophet_considered <- data.frame(
  ds = as.Date(paste0(trans_suicide$year, "-01-01")),
  y  = trans_suicide$considered_pct
)
prophet_attempted <- data.frame(
  ds = as.Date(paste0(trans_suicide$year, "-01-01")),
  y  = trans_suicide$attempted_pct
)

# Fit Prophet model
m_considered <- prophet(
  prophet_considered,
  yearly.seasonality = FALSE,
  weekly.seasonality = FALSE,
  daily.seasonality  = FALSE
)
```

```
n.changepoints greater than number of observations. Using 17
```

```r
# Fit Prophet model
m_attempted <- prophet(
  prophet_attempted,
  yearly.seasonality = FALSE,
  weekly.seasonality = FALSE,
  daily.seasonality  = FALSE
)
```

```
n.changepoints greater than number of observations. Using 17
```

```r
# Prophet variant with more flexible trend (larger changepoint prior)
m_considered_flex <- prophet(
  prophet_considered,
  yearly.seasonality = FALSE,
  weekly.seasonality = FALSE,
  daily.seasonality  = FALSE,
  changepoint.prior.scale = 0.5   # default is 0.05
)
```

```
n.changepoints greater than number of observations. Using 17
```

```r
m_attempted_flex <- prophet(
  prophet_attempted,
  yearly.seasonality = FALSE,
  weekly.seasonality = FALSE,
  daily.seasonality  = FALSE,
  changepoint.prior.scale = 0.5
)
```

```
n.changepoints greater than number of observations. Using 17
```

To evaluate robustness and guard against overfitting, time-series cross-validation was conducted for the Prophet models using rolling-origin evaluation. Models were repeatedly trained on expanding historical windows and evaluated on held-out future observations. This approach preserves temporal ordering and reflects the real-world task of learning from past data to explain or anticipate future trends. Cross-validation results were used comparatively to assess whether increased trend flexibility improved out-of-sample performance.

```r
# Cross-validation parameters
initial  <- 10*365
period   <- 1*365
horizon  <- 3*365
units <- "days"

# CV for baseline Prophet
cv_considered <- cross_validation(
  m_considered,
  initial  = initial,
  period   = period,
  horizon  = horizon,
  units = units
)
```

```
Making 10 forecasts with cutoffs between 2010-01-04 and 2019-01-02
```

```
n.changepoints greater than number of observations. Using 7
```

```
n.changepoints greater than number of observations. Using 8
```

```
n.changepoints greater than number of observations. Using 9
```

```
n.changepoints greater than number of observations. Using 10

n.changepoints greater than number of observations. Using 11
n.changepoints greater than number of observations. Using 11

n.changepoints greater than number of observations. Using 12

n.changepoints greater than number of observations. Using 13

n.changepoints greater than number of observations. Using 14

n.changepoints greater than number of observations. Using 15
```

```r
# CV for flexible changepoint Prophet
cv_considered_flex <- cross_validation(
  m_considered_flex,
  initial  = initial,
  period   = period,
  horizon  = horizon,
  units = units
)
```

```
Making 10 forecasts with cutoffs between 2010-01-04 and 2019-01-02

n.changepoints greater than number of observations. Using 7

n.changepoints greater than number of observations. Using 8

n.changepoints greater than number of observations. Using 9

n.changepoints greater than number of observations. Using 10

n.changepoints greater than number of observations. Using 11
n.changepoints greater than number of observations. Using 11

n.changepoints greater than number of observations. Using 12

n.changepoints greater than number of observations. Using 13
```

```
n.changepoints greater than number of observations. Using 14
```

```
n.changepoints greater than number of observations. Using 15
```

```r
# CV for baseline Prophet
cv_attempted <- cross_validation(
  m_attempted,
  initial  = initial,
  period   = period,
  horizon  = horizon,
  units = units
)
```

```
Making 10 forecasts with cutoffs between 2010-01-04 and 2019-01-02
```

```
n.changepoints greater than number of observations. Using 7
```

```
n.changepoints greater than number of observations. Using 8
```

```
n.changepoints greater than number of observations. Using 9
```

```
n.changepoints greater than number of observations. Using 10
```

```
n.changepoints greater than number of observations. Using 11
n.changepoints greater than number of observations. Using 11
```

```
n.changepoints greater than number of observations. Using 12
```

```
n.changepoints greater than number of observations. Using 13
```

```
n.changepoints greater than number of observations. Using 14
```

```
n.changepoints greater than number of observations. Using 15
```

```r
# CV for flexible changepoint Prophet
cv_attempted_flex <- cross_validation(
  m_attempted_flex,
  initial  = initial,
  period   = period,
  horizon  = horizon,
  units = units
)
```

```
Making 10 forecasts with cutoffs between 2010-01-04 and 2019-01-02
```

```
n.changepoints greater than number of observations. Using 7
```

```
n.changepoints greater than number of observations. Using 8
```

```
n.changepoints greater than number of observations. Using 9
```

```
n.changepoints greater than number of observations. Using 10
```

```
n.changepoints greater than number of observations. Using 11
n.changepoints greater than number of observations. Using 11
```

```
n.changepoints greater than number of observations. Using 12
```

```
n.changepoints greater than number of observations. Using 13
```

```
n.changepoints greater than number of observations. Using 14
```

```
n.changepoints greater than number of observations. Using 15
```

```r
perf_considered <- performance_metrics(cv_considered)
perf_considered_flex <- performance_metrics(cv_considered_flex)
perf_attempted <- performance_metrics(cv_attempted)
perf_attempted_flex <- performance_metrics(cv_attempted_flex)

cv_summary_considered <- bind_rows(
  perf_considered %>% mutate(model = "Prophet Considered Suicide (baseline"),
  perf_considered_flex %>% mutate(model = "Prophet Considered Suicide (flexible)")) %>%
```

```
  group_by(model) %>%
  summarise(
    MAE  = mean(mae),
    RMSE = mean(rmse),
    MAPE = mean(mape)
  )
cv_summary_attempted <- bind_rows(
  perf_attempted %>% mutate(model = "Prophet Attempted Suicide (baseline"),
  perf_attempted_flex %>% mutate(model = "Prophet Attempted Suicide (flexible)")
) %>%
  group_by(model) %>%
  summarise(
    MAE  = mean(mae),
    RMSE = mean(rmse),
    MAPE = mean(mape)
  )

cv_summary_considered
```

```
# A tibble: 2 x 4
  model                                   MAE  RMSE   MAPE
  <chr>                                 <dbl> <dbl>  <dbl>
1 Prophet Considered Suicide (baseline   2.74  3.09 0.0352
2 Prophet Considered Suicide (flexible)  3.14  3.86 0.0402
```

```
cv_summary_attempted
```

```
# A tibble: 2 x 4
  model                                  MAE  RMSE   MAPE
  <chr>                                <dbl> <dbl>  <dbl>
1 Prophet Attempted Suicide (baseline   2.37  2.61 0.0602
2 Prophet Attempted Suicide (flexible)  2.18  2.69 0.0545
```

```
# Forecast a few years ahead
future_considered <- make_future_dataframe(m_considered, periods = 3, freq = "year")
fc_considered <- predict(m_considered, future_considered)
# Forecast a few years ahead
future_attempted <- make_future_dataframe(m_attempted, periods = 3, freq = "year")
fc_attempted <- predict(m_attempted, future_attempted)
# Forecast a few years ahead
```

```r
future_considered_flex <- make_future_dataframe(
  m_considered_flex, periods = 3, freq = "year"
)
fc_considered_flex <- predict(m_considered_flex, future_considered_flex)
future_attempted_flex <- make_future_dataframe(
  m_attempted_flex, periods = 3, freq = "year"
)
fc_attempted_flex <- predict(m_attempted_flex, future_attempted_flex)

fc_considered[,c("ds", "trend", "yhat")]
```

```
           ds    trend     yhat
1  2000-01-01 61.51595 61.51595
2  2001-01-01 62.73800 62.73800
3  2002-01-01 63.95671 63.95671
4  2003-01-01 65.17542 65.17542
5  2004-01-01 66.39413 66.39413
6  2005-01-01 67.61617 67.61617
7  2006-01-01 68.83488 68.83488
8  2007-01-01 70.05359 70.05359
9  2008-01-01 71.27230 71.27230
10 2009-01-01 72.49435 72.49435
11 2010-01-01 73.71284 73.71284
12 2011-01-01 74.48086 74.48086
13 2012-01-01 75.24703 75.24703
14 2013-01-01 75.85191 75.85191
15 2014-01-01 76.45374 76.45374
16 2015-01-01 77.05291 77.05291
17 2016-01-01 77.65169 77.65169
18 2017-01-01 78.25212 78.25212
19 2018-01-01 78.85091 78.85091
20 2019-01-01 79.44969 79.44969
21 2020-01-01 80.04848 80.04848
22 2021-01-01 80.64891 80.64891
23 2022-01-01 81.24769 81.24769
24 2023-01-01 81.84648 81.84648
25 2024-01-01 82.44527 82.44527
26 2025-01-01 83.04569 83.04569
```

```r
fc_considered_flex[,c("ds", "trend", "yhat")]
```

```
           ds    trend     yhat
```

```
1  2000-01-01 61.54568 61.54568
2  2001-01-01 62.99535 62.99535
3  2002-01-01 64.44105 64.44105
4  2003-01-01 64.87080 64.87080
5  2004-01-01 65.30055 65.30055
6  2005-01-01 66.45928 66.45928
7  2006-01-01 67.61513 67.61513
8  2007-01-01 68.77097 68.77097
9  2008-01-01 70.87540 70.87540
10 2009-01-01 73.40707 73.40707
11 2010-01-01 75.93183 75.93183
12 2011-01-01 76.17660 76.17660
13 2012-01-01 76.42173 76.42173
14 2013-01-01 76.66753 76.66753
15 2014-01-01 76.91266 76.91266
16 2015-01-01 77.15590 77.15590
17 2016-01-01 77.39914 77.39914
18 2017-01-01 77.95374 77.95374
19 2018-01-01 78.50888 78.50888
20 2019-01-01 79.06402 79.06402
21 2020-01-01 79.61916 79.61916
22 2021-01-01 80.17583 80.17583
23 2022-01-01 80.73097 80.73097
24 2023-01-01 81.28611 81.28611
25 2024-01-01 81.84125 81.84125
26 2025-01-01 82.39791 82.39791
```

```r
fc_attempted[,c("ds", "trend", "yhat")]
```

```
           ds    trend     yhat
1  2000-01-01 29.11061 29.11061
2  2001-01-01 29.84446 29.84446
3  2002-01-01 30.57630 30.57630
4  2003-01-01 31.30814 31.30814
5  2004-01-01 32.03998 32.03998
6  2005-01-01 32.77383 32.77383
7  2006-01-01 33.50567 33.50567
8  2007-01-01 34.23751 34.23751
9  2008-01-01 34.96935 34.96935
10 2009-01-01 35.70301 35.70301
11 2010-01-01 36.43382 36.43382
12 2011-01-01 36.86331 36.86331
```

```
13 2012-01-01 37.29089 37.29089
14 2013-01-01 37.71889 37.71889
15 2014-01-01 38.14452 38.14452
16 2015-01-01 38.57013 38.57013
17 2016-01-01 38.99564 38.99564
18 2017-01-01 39.42232 39.42232
19 2018-01-01 39.84783 39.84783
20 2019-01-01 40.27335 40.27335
21 2020-01-01 40.69886 40.69886
22 2021-01-01 41.12553 41.12553
23 2022-01-01 41.55105 41.55105
24 2023-01-01 41.97656 41.97656
25 2024-01-01 42.40207 42.40207
26 2025-01-01 42.82875 42.82875
```

```r
fc_attempted_flex[,c("ds", "trend", "yhat")]
```

```
           ds    trend     yhat
1  2000-01-01 28.35469 28.35469
2  2001-01-01 29.19932 29.19932
3  2002-01-01 30.04165 30.04165
4  2003-01-01 30.88398 30.88398
5  2004-01-01 31.72631 31.72631
6  2005-01-01 32.57095 32.57095
7  2006-01-01 33.41335 33.41335
8  2007-01-01 34.25637 34.25637
9  2008-01-01 35.41247 35.41247
10 2009-01-01 36.57195 36.57195
11 2010-01-01 37.72826 37.72826
12 2011-01-01 37.93975 37.93975
13 2012-01-01 38.15112 38.15112
14 2013-01-01 38.36294 38.36294
15 2014-01-01 38.57411 38.57411
16 2015-01-01 38.78528 38.78528
17 2016-01-01 38.99762 38.99762
18 2017-01-01 39.28613 39.28613
19 2018-01-01 39.57385 39.57385
20 2019-01-01 39.86158 39.86158
21 2020-01-01 40.14930 40.14930
22 2021-01-01 40.43781 40.43781
23 2022-01-01 40.72553 40.72553
24 2023-01-01 41.01325 41.01325
```

```
25 2024-01-01 41.30097 41.30097
26 2025-01-01 41.58948 41.58948
```

Comparing Models

Model comparison was conducted separately for suicidal ideation (considered suicide) and suicide attempts using a combination of time-series cross-validation, information criteria, and diagnostic evaluation. Rolling-origin cross-validation was applied to the Prophet models for each outcome to assess out-of-sample performance while preserving the temporal structure of the data. For both considered and attempted suicide rates, the flexible-changepoint Prophet specification produced lower average forecast error than the baseline model, indicating that allowing greater adaptability in the estimated trend improved predictive performance. ARIMA and Holt models were evaluated using AICc and residual diagnostics for each series; although ARIMA residuals satisfied white-noise assumptions after differencing, the required differencing reduced interpretability of long-term trends that are central to the research question. Holt's linear trend models provided interpretable benchmarks for both outcomes but were less flexible in accommodating potential changes in trend slope. Taken together, results for both suicidal ideation and suicide attempts support the selection of the flexible-changepoint Prophet model as the final model, as it balances statistical performance, robustness, and substantive interpretability when assessing long-term changes in transgender suicide-related outcomes.

```r
AICc <- function(model) {
  k <- length(coef(model)) + 1  # parameters + variance
  n <- length(residuals(model))
  AIC(model) + (2 * k * (k + 1)) / (n - k - 1)
}

# ---- In-sample comparison for ARIMA and Holt models ----
model_ic_considered <- data.frame(
  model = c("ARIMA - Auto",
            "ARIMA - Manual",
            "Holt (undamped)",
            "Holt (damped)"),
  AICc = c(
    AICc(arima_considered_auto),
    AICc(arima_considered_manual),
    AICc(holt_considered$model),
    AICc(holt_damped_considered$model)
  )
)
model_ic_attempted <- data.frame(
  model = c("ARIMA - Auto",
```

```
            "ARIMA - Manual",
            "Holt (undamped)",
            "Holt (damped)"),
  AICc = c(
    AICc(arima_attempted_auto),
    AICc(arima_attempted_manual),
    AICc(holt_attempted$model),
    AICc(holt_damped_attempted$model)
  )
)

cv_summary_considered
```

```
# A tibble: 2 x 4
  model                                   MAE   RMSE   MAPE
  <chr>                                 <dbl>  <dbl>  <dbl>
1 Prophet Considered Suicide (baseline   2.74   3.09 0.0352
2 Prophet Considered Suicide (flexible)  3.14   3.86 0.0402
```

```
cv_summary_attempted
```

```
# A tibble: 2 x 4
  model                                  MAE   RMSE   MAPE
  <chr>                                <dbl>  <dbl>  <dbl>
1 Prophet Attempted Suicide (baseline   2.37   2.61 0.0602
2 Prophet Attempted Suicide (flexible)  2.18   2.69 0.0545
```

```
model_ic_considered
```

```
            model     AICc
1    ARIMA - Auto 107.2453
2  ARIMA - Manual 115.0304
3 Holt (undamped) 123.4415
4   Holt (damped) 121.4688
```

```
model_ic_attempted
```

```
            model      AICc
1    ARIMA - Auto  98.47215
2  ARIMA - Manual 106.17585
3 Holt (undamped) 114.71357
4   Holt (damped) 110.02636
```

Final Model and Conclusions

Final Model Specification

The final model selected for substantive interpretation, however, was the flexible-changepoint Prophet model. Prophet models the observed series as

$$y(t) = g(t) + \varepsilon(t),$$

where

$$g(t) = \left(k + \sum_{j=1}^{J} \delta_j \mathbf{1}(t \geq s_j)\right) t + \left(m - \sum_{j=1}^{J} \delta_j s_j \mathbf{1}(t \geq s_j)\right)$$

The terms of this expression include: - $y(t)$: observed percentage at year t - k: initial growth rate (slope) - m: initial intercept - s_j: estimated changepoint times - δ_j: change in slope at changepoint j - $\mathbf{1}(t \geq s_j)$: indicator function - $\varepsilon(t)$: error term

Because the data are annual and nonseasonal, seasonal components were omitted, and the model focuses entirely on estimating the long-run trend and its uncertainty on the original scale of the data. The relevant values to the prophet models may be observed below.

```r
# Extract core trend parameters
trend_params_considered <- tibble(
  parameter = c("initial_slope_k", "initial_intercept_m"),
  value = c(
    m_considered_flex$params$k,
    m_considered_flex$params$m
  )
)
trend_params_attempted <- tibble(
  parameter = c("initial_slope_k", "initial_intercept_m"),
  value = c(
    m_attempted_flex$params$k,
    m_attempted_flex$params$m
  )
)
changepoints_considered <- tibble(
  changepoint_date = m_considered_flex$changepoints,
  delta_slope = as.numeric(m_considered_flex$params$delta)
)
changepoints_attempted <- tibble(
  changepoint_date = m_attempted_flex$changepoints,
  delta_slope = as.numeric(m_attempted_flex$params$delta)
)
```

```
fitted_trend_considered <- fc_considered_flex |>
  select(ds, trend, yhat)
fitted_trend_attempted <- fc_attempted_flex |>
  select(ds, trend, yhat)

fitted_trend_considered
```

```
           ds    trend     yhat
1  2000-01-01 61.54568 61.54568
2  2001-01-01 62.99535 62.99535
3  2002-01-01 64.44105 64.44105
4  2003-01-01 64.87080 64.87080
5  2004-01-01 65.30055 65.30055
6  2005-01-01 66.45928 66.45928
7  2006-01-01 67.61513 67.61513
8  2007-01-01 68.77097 68.77097
9  2008-01-01 70.87540 70.87540
10 2009-01-01 73.40707 73.40707
11 2010-01-01 75.93183 75.93183
12 2011-01-01 76.17660 76.17660
13 2012-01-01 76.42173 76.42173
14 2013-01-01 76.66753 76.66753
15 2014-01-01 76.91266 76.91266
16 2015-01-01 77.15590 77.15590
17 2016-01-01 77.39914 77.39914
18 2017-01-01 77.95374 77.95374
19 2018-01-01 78.50888 78.50888
20 2019-01-01 79.06402 79.06402
21 2020-01-01 79.61916 79.61916
22 2021-01-01 80.17583 80.17583
23 2022-01-01 80.73097 80.73097
24 2023-01-01 81.28611 81.28611
25 2024-01-01 81.84125 81.84125
26 2025-01-01 82.39791 82.39791
```

```
changepoints_considered
```

```
# A tibble: 17 x 2
   changepoint_date    delta_slope
   <dttm>                    <dbl>
 1 2001-01-01 00:00:00    -2.31e-8
```

```
 2 2002-01-01 00:00:00    -2.69e-1
 3 2003-01-01 00:00:00     1.70e-8
 4 2004-01-01 00:00:00     1.93e-1
 5 2005-01-01 00:00:00     7.42e-5
 6 2006-01-01 00:00:00    -1.66e-8
 7 2007-01-01 00:00:00     2.52e-1
 8 2008-01-01 00:00:00     1.11e-1
 9 2009-01-01 00:00:00     3.92e-9
10 2010-01-01 00:00:00    -6.05e-1
11 2011-01-01 00:00:00     9.62e-5
12 2012-01-01 00:00:00    -1.03e-8
13 2013-01-01 00:00:00     2.55e-8
14 2014-01-01 00:00:00    -5.01e-4
15 2015-01-01 00:00:00    -7.68e-9
16 2016-01-01 00:00:00     8.22e-2
17 2017-01-01 00:00:00     5.48e-4
```

```
trend_params_considered
```

```
# A tibble: 2 x 2
  parameter           value
  <chr>               <dbl>
1 initial_slope_k     0.383
2 initial_intercept_m 0.742
```

```
fitted_trend_attempted
```

```
           ds    trend     yhat
1  2000-01-01 28.35469 28.35469
2  2001-01-01 29.19932 29.19932
3  2002-01-01 30.04165 30.04165
4  2003-01-01 30.88398 30.88398
5  2004-01-01 31.72631 31.72631
6  2005-01-01 32.57095 32.57095
7  2006-01-01 33.41335 33.41335
8  2007-01-01 34.25637 34.25637
9  2008-01-01 35.41247 35.41247
10 2009-01-01 36.57195 36.57195
11 2010-01-01 37.72826 37.72826
12 2011-01-01 37.93975 37.93975
13 2012-01-01 38.15112 38.15112
```

```
14 2013-01-01 38.36294 38.36294
15 2014-01-01 38.57411 38.57411
16 2015-01-01 38.78528 38.78528
17 2016-01-01 38.99762 38.99762
18 2017-01-01 39.28613 39.28613
19 2018-01-01 39.57385 39.57385
20 2019-01-01 39.86158 39.86158
21 2020-01-01 40.14930 40.14930
22 2021-01-01 40.43781 40.43781
23 2022-01-01 40.72553 40.72553
24 2023-01-01 41.01325 41.01325
25 2024-01-01 41.30097 41.30097
26 2025-01-01 41.58948 41.58948
```

```
changepoints_attempted
```

```
# A tibble: 17 x 2
   changepoint_date    delta_slope
   <dttm>                    <dbl>
 1 2001-01-01 00:00:00    -6.00e-9
 2 2002-01-01 00:00:00    -1.05e-9
 3 2003-01-01 00:00:00     4.53e-8
 4 2004-01-01 00:00:00     2.57e-8
 5 2005-01-01 00:00:00     3.18e-5
 6 2006-01-01 00:00:00     3.16e-4
 7 2007-01-01 00:00:00     1.57e-1
 8 2008-01-01 00:00:00     1.08e-4
 9 2009-01-01 00:00:00     1.16e-7
10 2010-01-01 00:00:00    -4.73e-1
11 2011-01-01 00:00:00    -6.28e-5
12 2012-01-01 00:00:00    -6.46e-5
13 2013-01-01 00:00:00    -3.31e-5
14 2014-01-01 00:00:00     8.75e-8
15 2015-01-01 00:00:00     5.82e-4
16 2016-01-01 00:00:00     3.77e-2
17 2017-01-01 00:00:00    -1.07e-8
```

```
trend_params_attempted
```

```
# A tibble: 2 x 2
  parameter           value
```

```
  <chr>               <dbl>
1 initial_slope_k     0.421
2 initial_intercept_m 0.644
```

Answering the Research Question

Statistically, results from the Prophet model indicate a clear and sustained upward trend in both suicidal ideation and suicide attempts among transgender individuals in the United States from 2000 to 2022. For suicidal ideation, the estimated trend increased from approximately 61–63% in the early 2000s to point forecasts exceeding 80% by the most recent years, with 95% prediction intervals in the low-to-mid 80% range. For suicide attempts, the trend increased from approximately 28–30% in the early 2000s to point forecasts above 40% in recent years, with corresponding prediction intervals spanning the high 30s to low 40s. The fact that prediction intervals from both Prophet variants consistently exclude earlier historical levels provides strong statistical evidence that recent rates represent a meaningful increase rather than random fluctuation.

In practical terms, these findings suggest that elevated suicide risk among transgender individuals is not a transient phenomenon but reflects a long-term escalation over more than two decades. This has direct implications for public health planning and policy: interventions must be sustained, structural, and responsive to long-run social conditions rather than short-term crises alone. The results also underscore the urgency of addressing systemic contributors such as discrimination, barriers to affirming healthcare, and social marginalization.

Strengths and Limitations of ARIMA Relative to Prophet

ARIMA models were useful for assessing whether residual temporal dependence remained after detrending and for validating that the series exhibit nonstationarity driven by strong trends. However, the need for differencing substantially limits interpretability with respect to long-term growth, which is central to the research question. Moreover, ARIMA assumes stable data-generating processes over time and is less well suited to capturing gradual accelerations or structural changes in trend.

Prophet overcomes these limitations by modeling trends directly on the original scale of the data and allowing for changepoints that reflect potential shifts in underlying processes. Its flexibility and interpretability make it particularly valuable for this application, though its conclusions remain constrained by the short length of the annual series and reliance on aggregated survey estimates. Importantly, Prophet does not eliminate uncertainty about causality or mechanism; it provides a clearer descriptive and predictive account of long-run change rather than a causal explanation.

Final Conclusions and Generalizability

Using a combination of exploratory analysis, ARIMA diagnostics, Holt-based trend models, and cross-validated Prophet models, this study finds strong evidence that both suicidal ideation and suicide attempt rates among transgender individuals in the United States have increased substantially from 2000 to 2022. The final conclusions rely primarily on the flexible-changepoint Prophet model, supported by consistency across alternative modeling approaches rather than any single estimate. While the results generalize most directly to the U.S. transgender population as represented in national survey data, they may not fully capture heterogeneity across subgroups or changes in reporting practices over time.

Overall, the analysis demonstrates that long-term trend-focused models provide critical insight that would be obscured by purely short-memory approaches. The findings highlight the persistence and magnitude of suicide-related disparities and reinforce the need for long-term, evidence-based policy responses rather than episodic interventions.