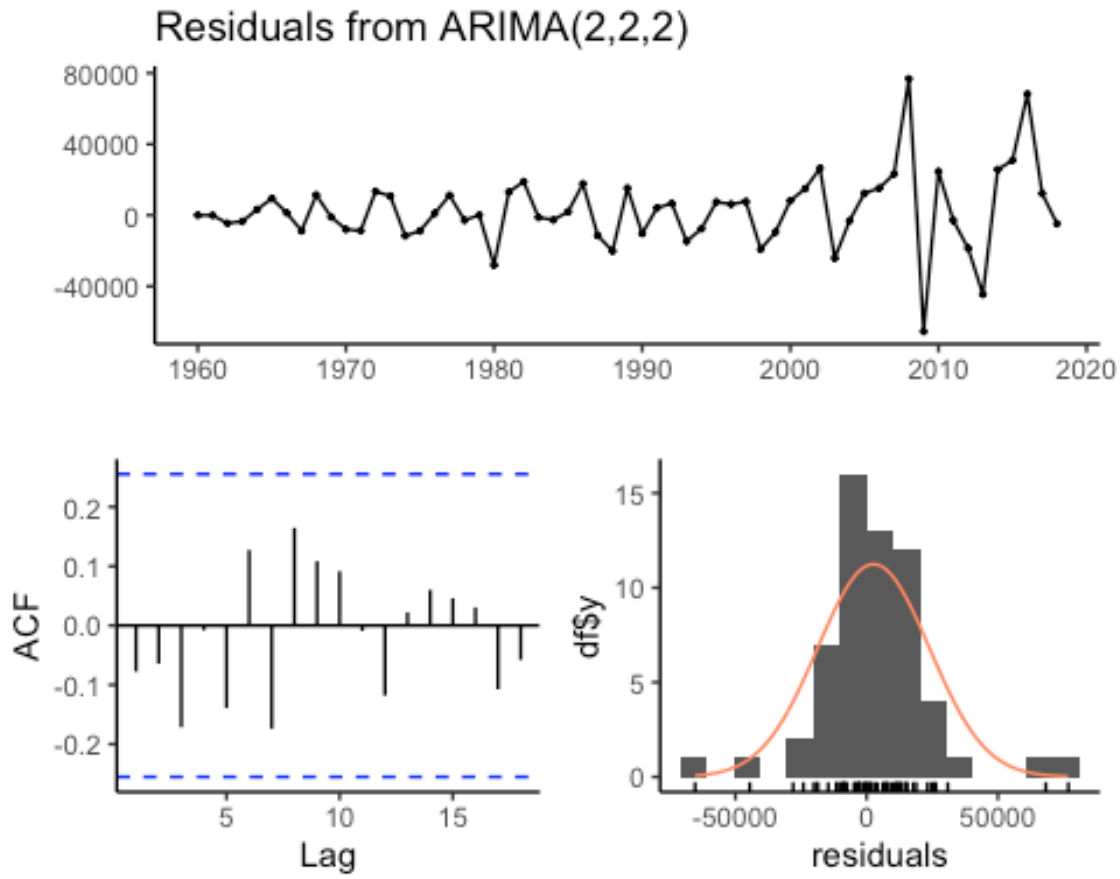
DSCI 5340 Extra Credit: Use ChatGPT to Create a Forecast Model Satyasriharsha Buddha

**Goal:** Using Forecasting, we want to know usage of Coal in the next 3 years

**Data:** For this model, we use this dataset from Github.

**Data Description:**
Time: The time variable represents the year in chronological order.
Usage: The usage variable represents the energy consumption in each year.

```
pacman::p_load(fpp3, lubridate, tidyverse)
theme_set(theme_classic())

# Load the forecast package
library(forecast)

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
## Attaching package: 'forecast'

## The following object is masked from 'package:fabletools':
##     accuracy

# Read the data from the CSV file
data <- read.csv("Downloads/energy_consumption.csv")

# Create a time series object
consumption <- ts(data$Consumption, start = c(1960), frequency = 1)

# Fit the ARIMA model
model <- auto.arima(consumption)

# Forecast the next 3 years
forecast <- forecast(model, h = 3)

# Print the forecasted values
print(forecast)

##      Point Forecast     Lo 80   Hi 80     Lo 95   Hi 95
## 2019        1026131  997400.7 1054862  982191.6 1070071
## 2020        1119506 1070207.7 1168805 1044110.6 1194902
## 2021        1219365 1141794.2 1296935 1100730.9 1337998

# Residual diagnostics
checkresiduals(model)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,2,2)
## Q* = 10.401, df = 6, p-value = 0.1088
##
## Model df: 4.   Total lags used: 10
```

```
# Calculate prediction intervals
prediction_interval <- forecast(forecast, level = c(80, 95))

# Print the prediction intervals
print(prediction_interval)
```

```
##      Point Forecast     Lo 80   Hi 80     Lo 95   Hi 95
## 2019        1026131  997400.7 1054862  982191.6 1070071
## 2020        1119506 1070207.7 1168805 1044110.6 1194902
## 2021        1219365 1141794.2 1296935 1100730.9 1337998
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

How can you improve the forecasting model generated by ChatGPT:

(i) Coding Improvement:

The initial code lacked some important features. Data preprocessing steps were missing, such as handling missing values and outliers. Additionally, there was no split between training and testing data, which is necessary to evaluate the model's performance.

(ii) Considering Better Models:

While the ARIMA model is commonly used for time series forecasting, the code did not explore alternative models that might have been more suitable for the given data. Alternative approaches like exponential smoothing (e.g., Holt-Winters) could have been evaluated to improve forecasting accuracy. The code also lacked model selection validation based on statistical criteria or cross-validation techniques.