Time Series Project

AnnaMarie Fernandes and Ryan Schaefer

Import Dependencies

```
library(astsa)
library(tidyverse)
library(tswge)
library(tseries)
library(forecast)
library(knitr)
library(Metrics)
```

Import Data

```
# nfl = read.csv("nfl_attendance.csv")
# save(nfl, file = "nfl_attendance.RData")
load("nfl_attendance.RData")

head(nfl)
```

```
     team team_name year  total   home   away week weekly_attendance
1 Arizona Cardinals 2000 893926 387475 506451    1             77434
2 Arizona Cardinals 2000 893926 387475 506451    2             66009
3 Arizona Cardinals 2000 893926 387475 506451    3                NA
4 Arizona Cardinals 2000 893926 387475 506451    4             71801
5 Arizona Cardinals 2000 893926 387475 506451    5             66985
6 Arizona Cardinals 2000 893926 387475 506451    6             44296
```

Format Data

```
nfl.cumulative = nfl %>%
  group_by(year, week) %>%
  summarize(attendance = sum(weekly_attendance, na.rm = TRUE)) %>%
  ungroup()
```

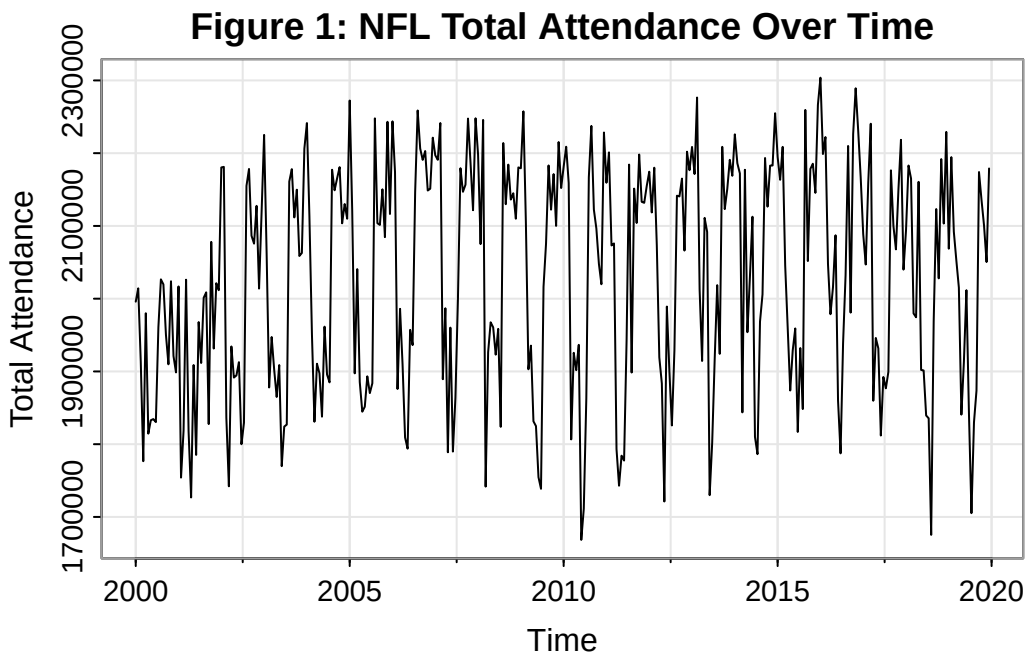
```
`summarise()` has grouped output by 'year'. You can override using the
`.groups` argument.
```

```
head(nfl.cumulative)
```

```
# A tibble: 6 x 3
   year  week attendance
  <int> <int>      <int>
1  2000     1    1995898
2  2000     2    2014246
3  2000     3    1913296
4  2000     4    1776618
5  2000     5    1980014
6  2000     6    1814476
```

Plot Time Series

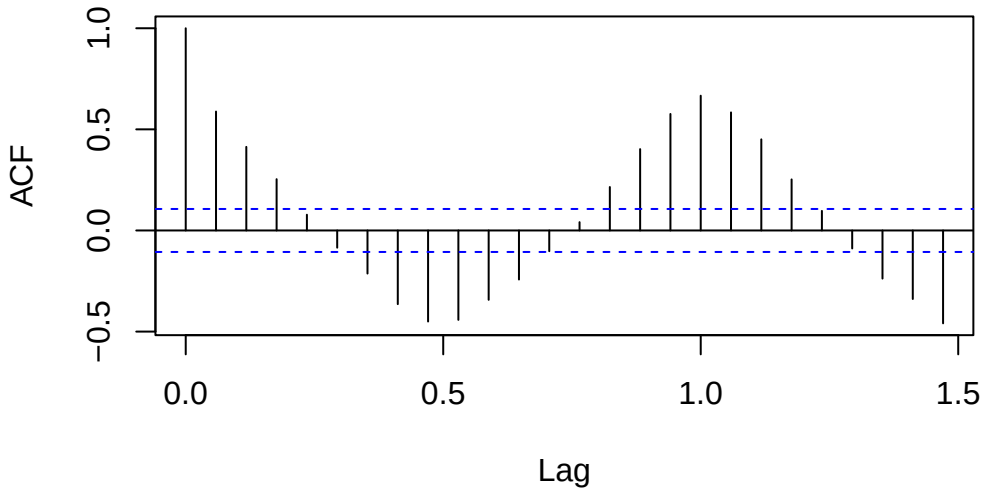
```
nfl.cumulative.ts = ts(nfl.cumulative$attendance, start = c(2000, 1), frequency = 17)

tsplot(
  nfl.cumulative.ts,
  main = "Figure 1: NFL Total Attendance Over Time",
  ylab = "Total Attendance"
)
```



ACF, PACF and Spectral Density

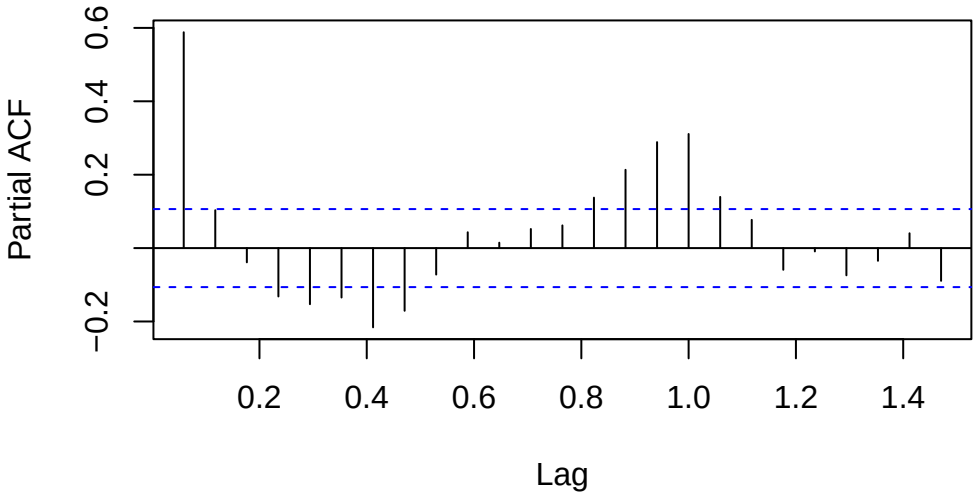
```
acf(
  nfl.cumulative.ts,
  main = "Figure 2: NFL Total Attendance ACF"
)
```

Figure 2: NFL Total Attendance ACF



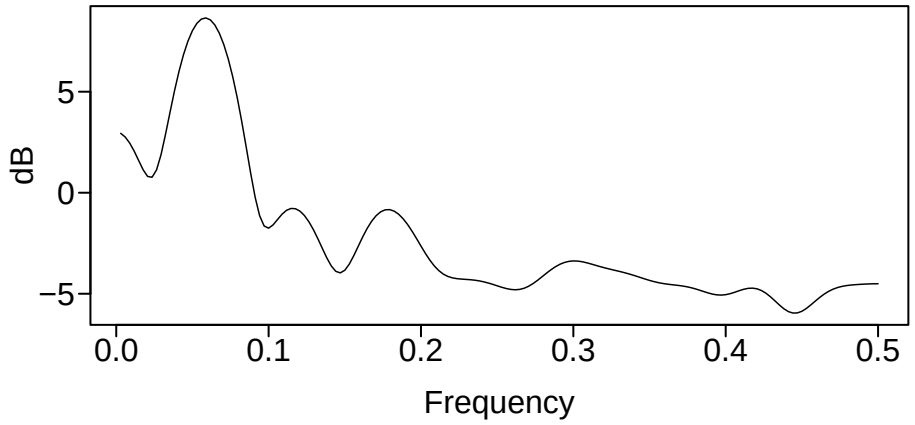

```
pacf(
  nfl.cumulative.ts,
  main = "Figure 3: NFL Total Attendance PACF"
)
```

Figure 3: NFL Total Attendance PACF




```
specDensity = parzen.wge(nfl.cumulative.ts)
```

Parzen Window Truncation point: M = 36



Check for stationarity

```
adf.test(nfl.cumulative.ts)
```

```
Warning in adf.test(nfl.cumulative.ts): p-value smaller than printed p-value
```

```

    Augmented Dickey-Fuller Test

data:  nfl.cumulative.ts
Dickey-Fuller = -10.112, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

Training/Testing Split

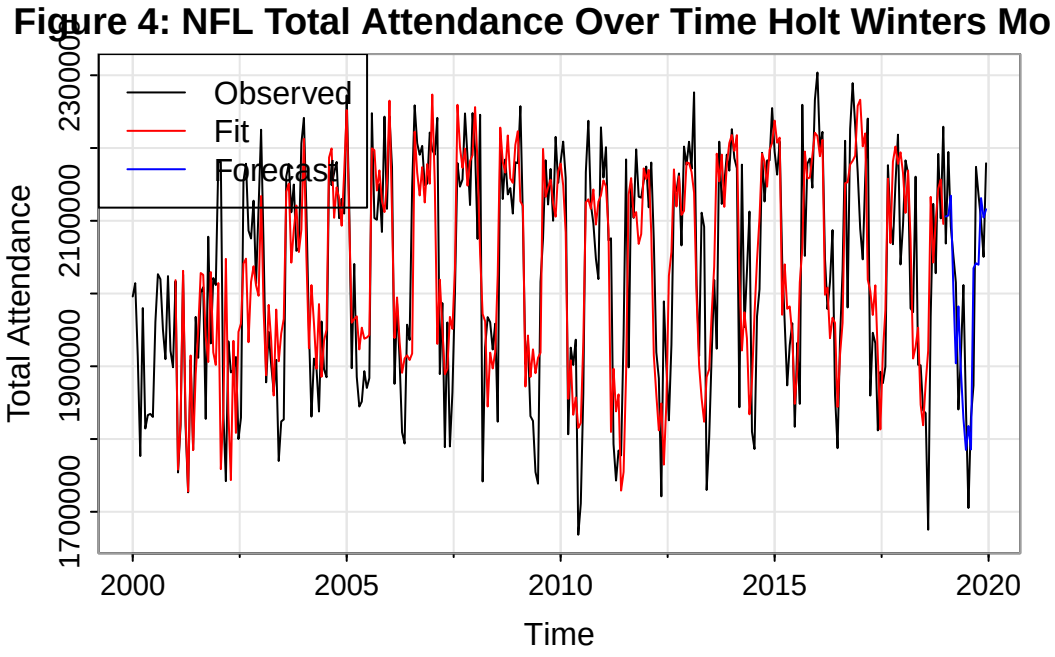
```
nfl.training = nfl.cumulative %>%
  dplyr::filter(year < 2019)

nfl.training.ts = ts(nfl.training$attendance, start = c(2000, 1), frequency = 17)

nfl.testing = nfl.cumulative %>%
  dplyr::filter(year == 2019)

nfl.testing.ts = ts(nfl.testing$attendance, start = c(2019, 1), frequency = 17)
```

Holt Winters Model

Fit Model

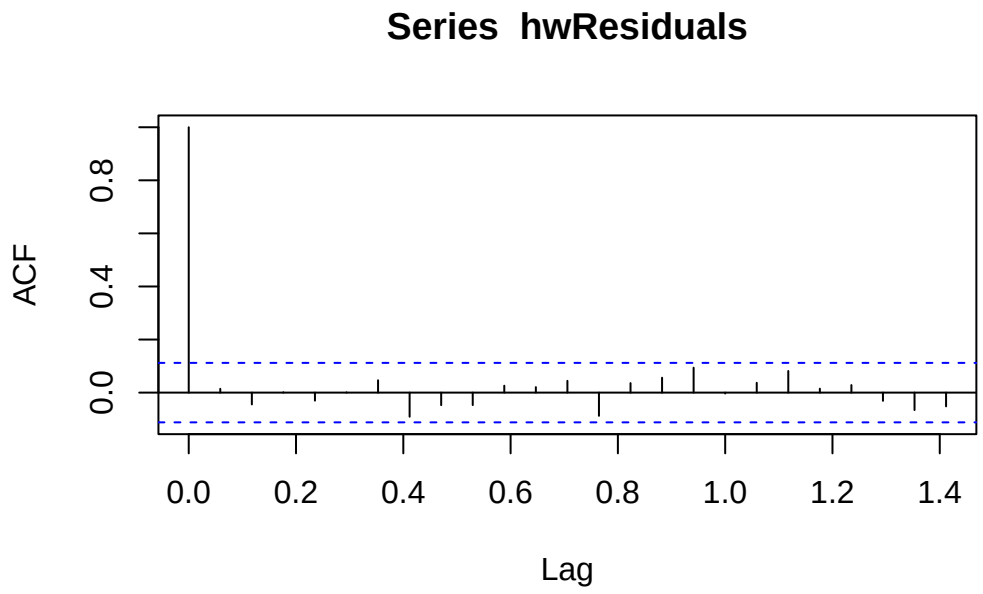
```
hwModel = HoltWinters(nfl.training.ts)
hwForecast = forecast(
  hwModel,
  h = 17
)

tsplot(
  nfl.cumulative.ts,
```

```
  main = "Figure 4: NFL Total Attendance Over Time Holt Winters Model",
  ylab = "Total Attendance"
)
lines(hwForecast$fitted, col = "red")
lines(hwForecast$mean, col = "blue")

legend(
  "topleft",
  legend = c("Observed", "Fit", "Forecast"),
  lty = 1,
  col = c("black", "red", "blue")
)
```



Model Evaluation

```
hwResiduals = window(nfl.training.ts, start = 2001) - hwForecast$fitted
acf(hwResiduals)
```




```
Box.test(hwResiduals)
```

```
    Box-Pierce test

data:  hwResiduals
X-squared = 0.063451, df = 1, p-value = 0.8011
```

```
rmse.hw = rmse(nfl.testing$attendance, hwForecast$mean)
rmse.hw
```

```
[1] 96414.18
```

SARIMA

Determine Model Orders

```
auto.arima(nfl.training.ts, approximation = FALSE)
```

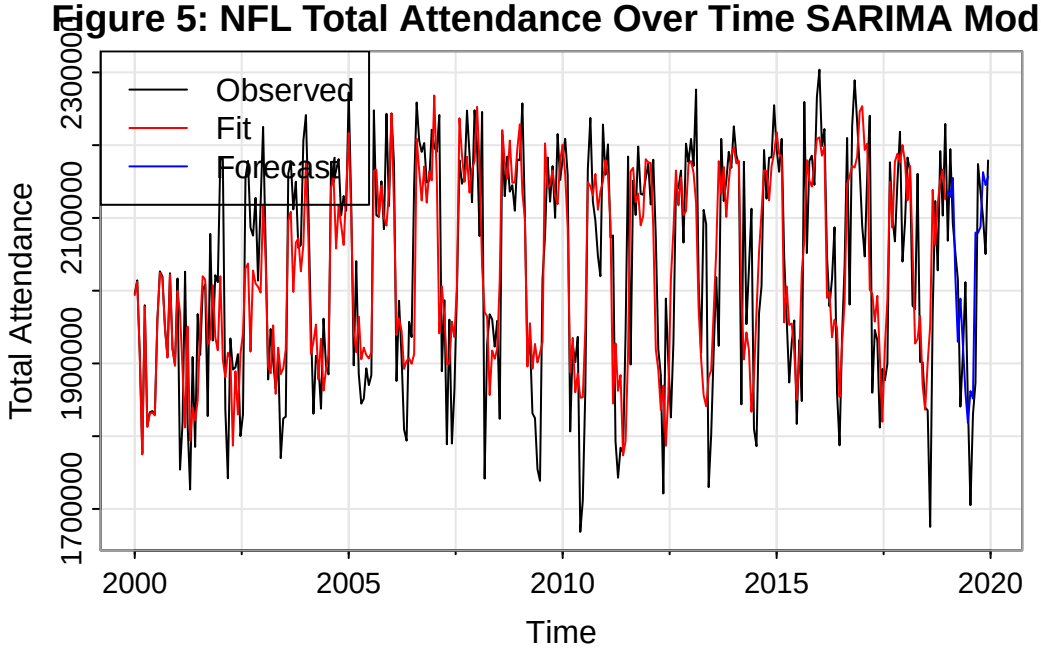
```
Series: nfl.training.ts
ARIMA(1,0,1)(0,1,1)[17] with drift

Coefficients:
         ar1      ma1     sma1     drift
      0.9857  -0.9412  -0.6985  326.3695
s.e.  0.0170   0.0260   0.0466  396.4345

sigma^2 = 9.884e+09:  log likelihood = -3958.77
AIC=7927.55   AICc=7927.75   BIC=7946.17
```

Fit Model

```
sarimaModel = arima(nfl.training.ts, order = c(1, 0, 1), seasonal = list(order = c(0, 1, 1
sarimaForecast = predict(sarimaModel, n.ahead = 17)
```

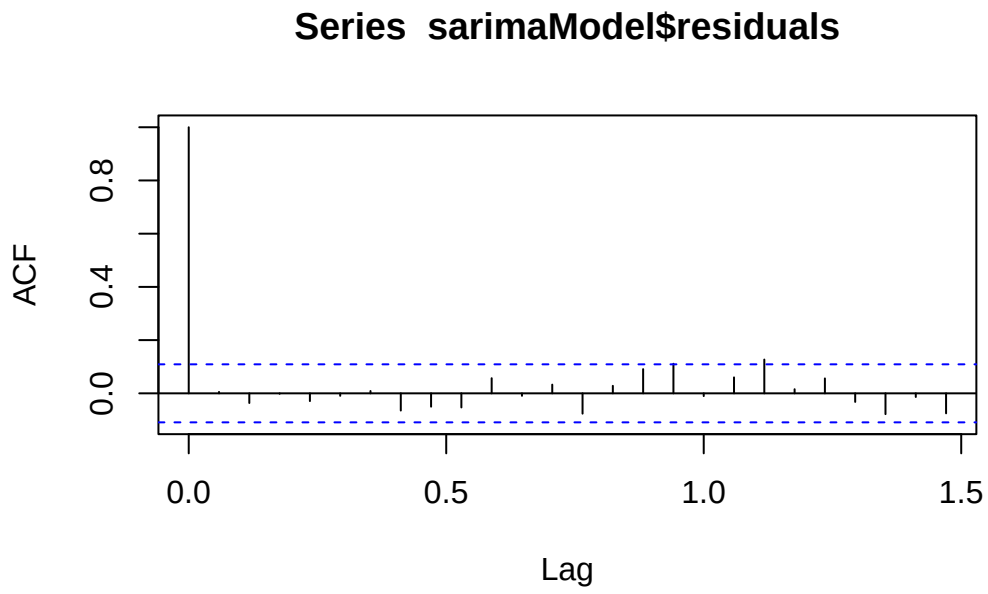
```
sarimaFit = nfl.training.ts - sarimaModel$residuals

tsplot(
  nfl.cumulative.ts,
  main = "Figure 5: NFL Total Attendance Over Time SARIMA Model",
  ylab = "Total Attendance"
)
lines(sarimaFit, col = "red")
lines(sarimaForecast$pred, col = "blue")

legend(
  "topleft",
  legend = c("Observed", "Fit", "Forecast"),
  lty = 1,
  col = c("black", "red", "blue")
)
```

Figure 5: NFL Total Attendance Over Time SARIMA Mode



Model Evaluation

```
acf(sarimaModel$residuals)
```




```
Box.test(sarimaModel$residuals)
```

```
    Box-Pierce test

data:  sarimaModel$residuals
X-squared = 0.0096764, df = 1, p-value = 0.9216
```

```
rmse.sarima = rmse(nfl.testing$attendance, sarimaForecast$pred)
rmse.sarima
```

```
[1] 96660.92
```

Linear Regression

Fit Model

```
nfl.cumulative2 = nfl.cumulative[18:nrow(nfl.cumulative),] %>%
  mutate(
    t = 18:nrow(nfl.cumulative),
    sin17 = sin((2 * pi * t) / 17),
    cos17 = cos((2 * pi * t) / 17),
    attendance.lag = nfl.cumulative$attendance[t - 17]
  )

head(nfl.cumulative2)
```

```
# A tibble: 6 x 7
   year  week attendance     t sin17   cos17 attendance.lag
  <int> <int>      <int> <int> <dbl>   <dbl>          <int>
1  2001     1    2016864    18 0.361  0.932         1995898
2  2001     2    1754008    19 0.674  0.739         2014246
3  2001     3    1818632    20 0.895  0.446         1913296
4  2001     4    2026086    21 0.996  0.0923        1776618
5  2001     5    1817838    22 0.962 -0.274         1980014
6  2001     6    1726716    23 0.798 -0.603         1814476
```

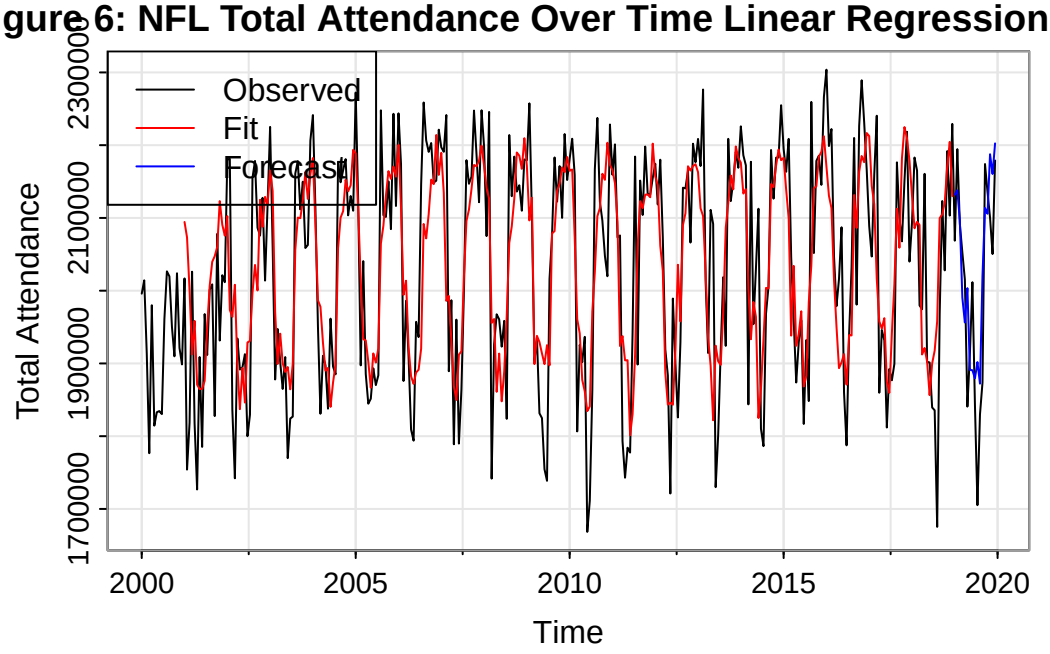
```
nfl.cumulative2.training = nfl.cumulative2 %>%
  filter(year < 2019)

nfl.cumulative2.testing = nfl.cumulative2 %>%
  filter(year == 2019)

lmModel = lm(attendance ~ sin17 + cos17 + attendance.lag, data = nfl.cumulative2.training)

summary(lmModel)
```

```
Call:
lm(formula = attendance ~ sin17 + cos17 + attendance.lag, data = nfl.cumulative2.training)

Residuals:
    Min      1Q  Median      3Q     Max
```

```
-319227  -63226    1791   56355  287738

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.257e+06  1.061e+05  11.845  < 2e-16 ***
sin17          -3.430e+04  8.699e+03  -3.943 9.99e-05 ***
cos17           9.063e+04  1.071e+04   8.459 1.18e-15 ***
attendance.lag  3.835e-01  5.220e-02   7.347 1.90e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 99220 on 302 degrees of freedom
Multiple R-squared:  0.5847,    Adjusted R-squared:  0.5805
F-statistic: 141.7 on 3 and 302 DF,  p-value: < 2.2e-16
```

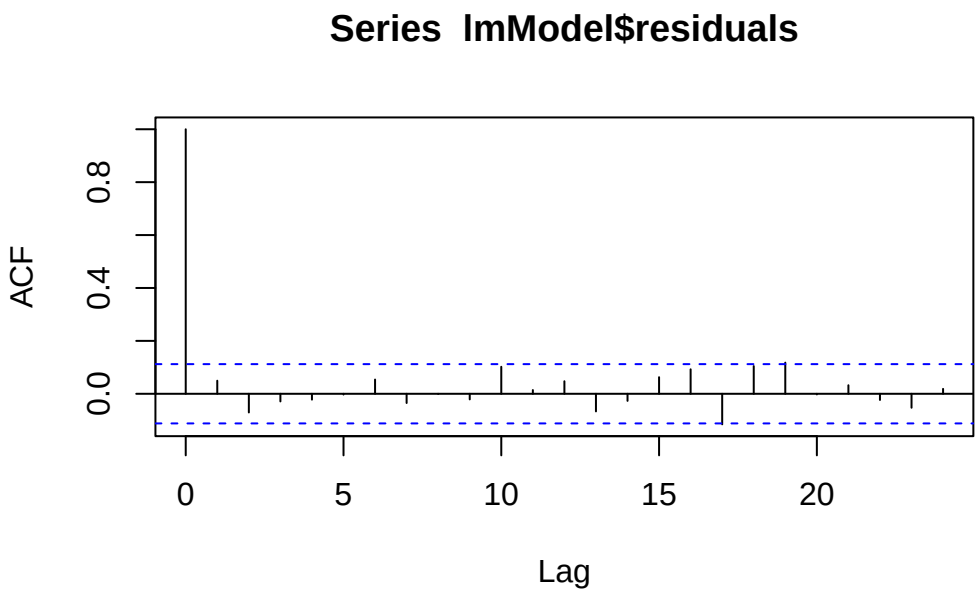
```r
lmModel.predict = predict(lmModel, newdata = nfl.cumulative2.testing)

ts.lmModel = ts(lmModel$fitted.values, start = c(2001, 1), frequency = 17)
ts.lmPredict = ts(lmModel.predict, start = c(2019, 1), frequency = 17)

tsplot(
  nfl.cumulative.ts,
  main = "Figure 6: NFL Total Attendance Over Time Linear Regression Model",
  ylab = "Total Attendance"
)
lines(ts.lmModel, col = "red")
lines(ts.lmPredict, col = "blue")

legend(
  "topleft",
  legend = c("Observed", "Fit", "Forecast"),
  lty = 1,
  col = c("black", "red", "blue")
)
```

gure 6: NFL Total Attendance Over Time Linear Regression



Model Evalutation

```
acf(lmModel$residuals)
```




```
Box.test(lmModel$residuals)
```

```
    Box-Pierce test

data:  lmModel$residuals
X-squared = 0.75023, df = 1, p-value = 0.3864
```

```
rmse.lm = rmse(nfl.cumulative2.testing$attendance, lmModel.predict)
rmse.lm
```

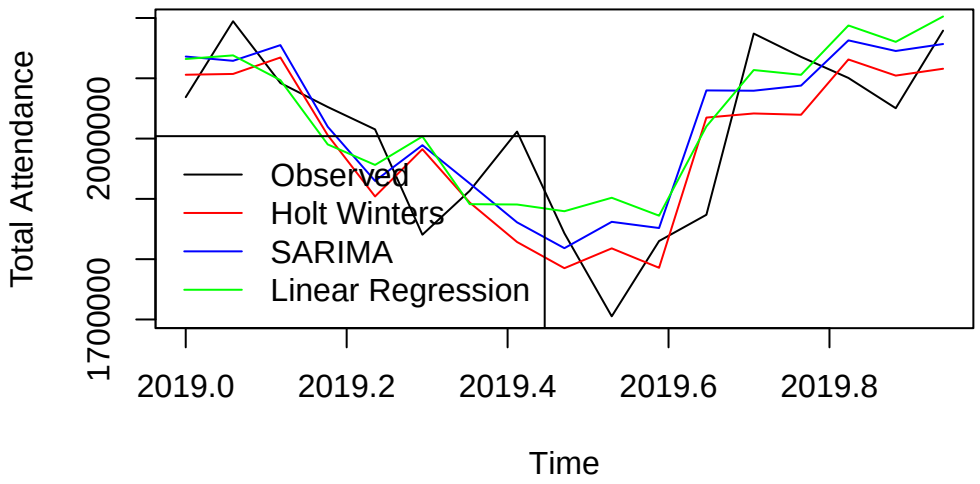
```
[1] 92131.05
```

Final Results

```
plot(
  nfl.testing.ts,
  main = "Figure 7: 2019 NFL Total Attendance vs Model Forecasts",
```

```
  ylab = "Total Attendance"
)

lines(hwForecast$mean, col = "red")
lines(sarimaForecast$pred, col = "blue")
lines(ts.lmPredict, col = "green")

legend(
  "bottomleft",
  legend = c(
    "Observed",
    "Holt Winters",
    "SARIMA",
    "Linear Regression"
  ),
  lty = 1,
  col = c(
    "black",
    "red",
    "blue",
    "green"
  )
)
```

Figure 7: 2019 NFL Total Attendance vs Model Forecasts




```
data.frame(
  model = c(
    "Holt Winters",
    "SARIMA",
    "Linear Regression"
  ),
  rmse = c(
    rmse.hw,
    rmse.sarima,
    rmse.lm
  )
) %>%
  kable(
    col.names = c(
      "Model Type",
      "RMSE"
    ),
    caption = "RMSE of NFL Attendance Forecasts"
  )
```

Table 1: RMSE of NFL Attendance Forecasts

Model Type	RMSE
Holt Winters	96414.18
SARIMA	96660.92
Linear Regression	92131.05