

# Startups Funding

Michael Audie, Karen Thornton, Hannah Schofield

2023-04-28

**Background** + Clean Tech startups are on the rise + There are two stages of funding stages: early stage and late stage. Early stage can be considered seed funding and late stage can be something like series C funding. Late stage funding is often more money for fewer companies + Companies that make it to late stage funding have a higher chance of being successful. + Greater late stage funding investment suggests a more maturation of the Clean Tech industry.

## Research Questions

1. Are investments in clean tech startups expected to continue to increase?
2. As the clean tech industry matures will investments in late-stage startups outpace investments in earlier-stage startups?

**Data** The data was collected from the IEA. The data is broken up into early and late stage investments. The data original went to 2022, but with Covid, the War in Ukraine, and a recession, there were dips in investment in both stages. Since we believe this doesn't reflect the trend of the industry, we decided to remove the year 2022.

**Limitations** The dataset had some limitations that made it harder to predict future years. These limitations include, but are not limited to: 1. The data was yearly, as opposed to daily or monthly. This made it difficult to show seasonality. It also made the data less granular and with less data points it was harder to see a trend to model. 2. As mentioned above, covid, and other factors like the war in Ukraine, and overall recession could be factors in the model that cannot be predicted or anticipated.

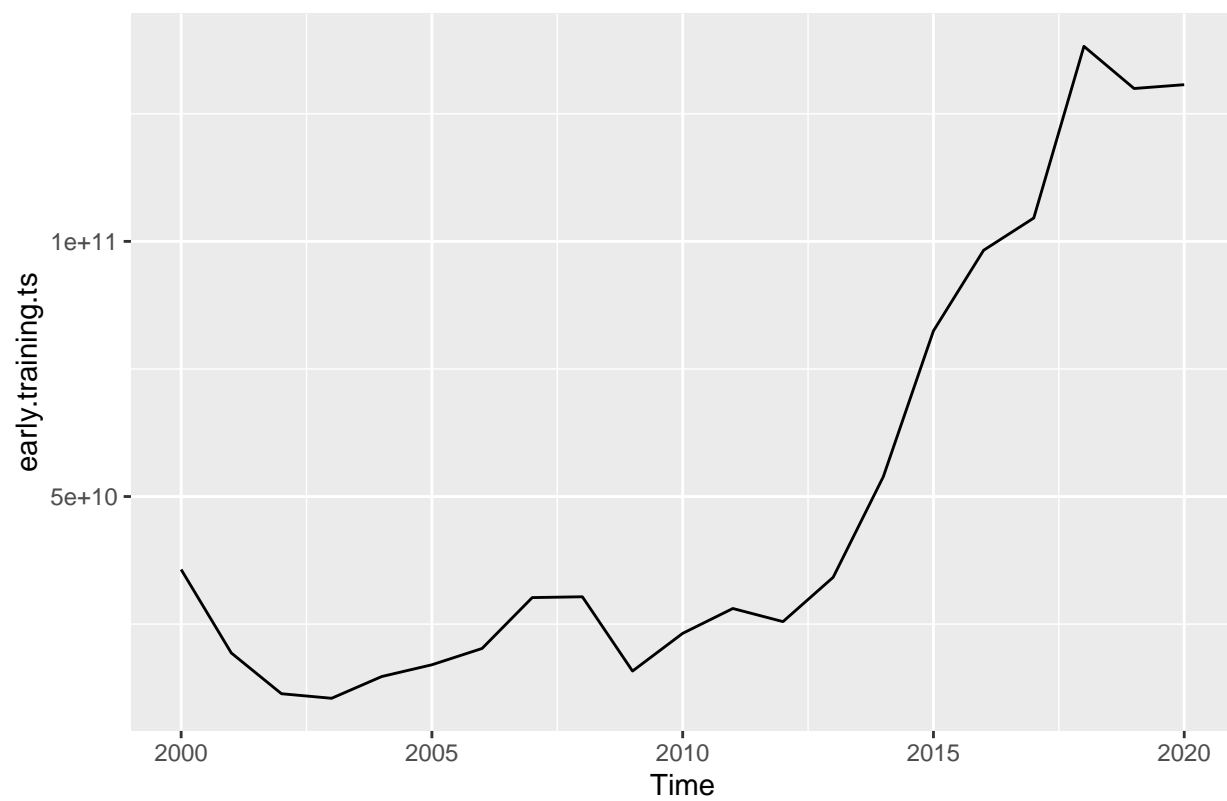
## Data Processing Steps

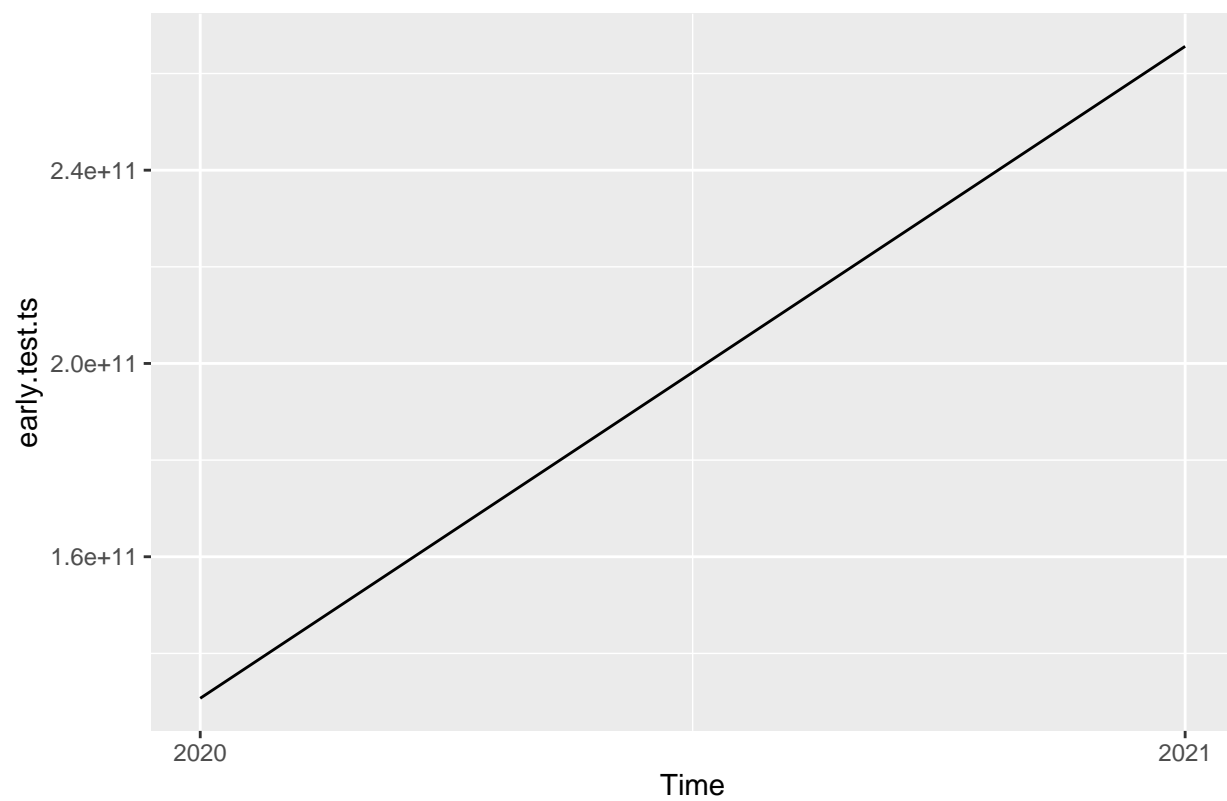
#Data Wrangling - Full sets

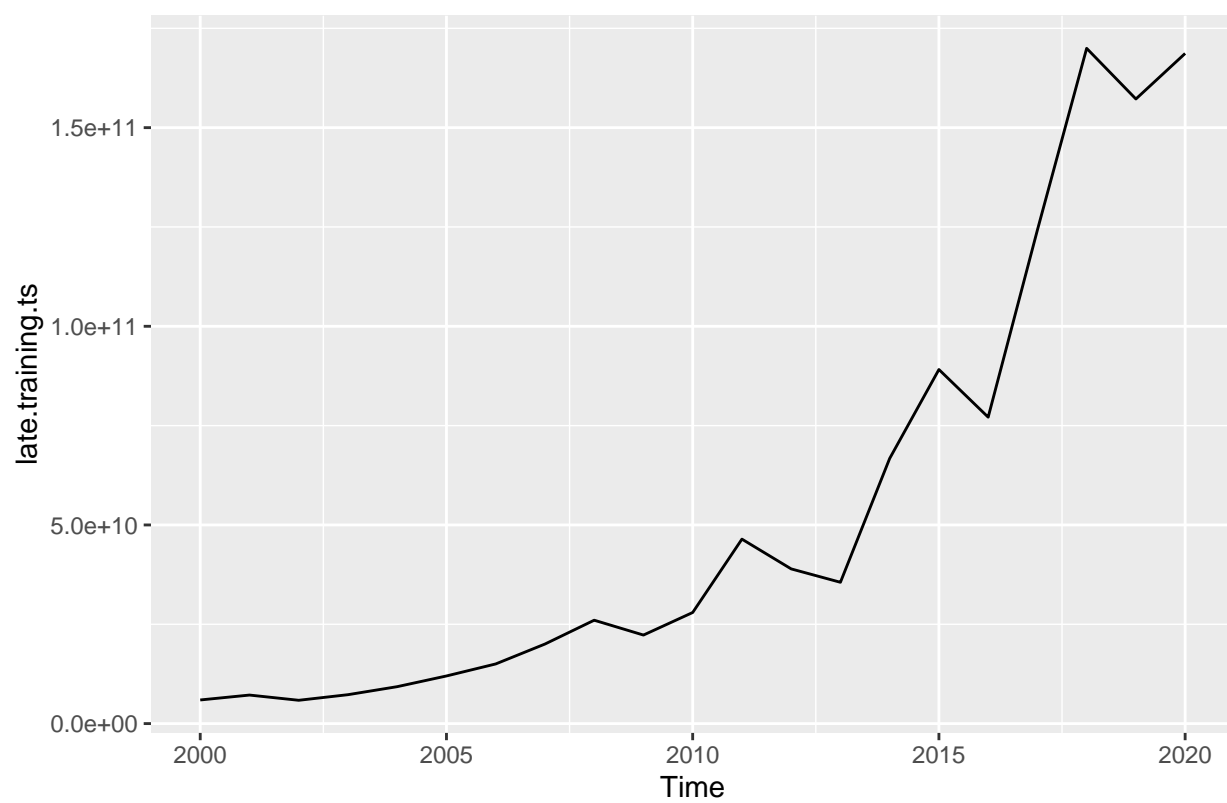
#Data Wrangling - Training Sets

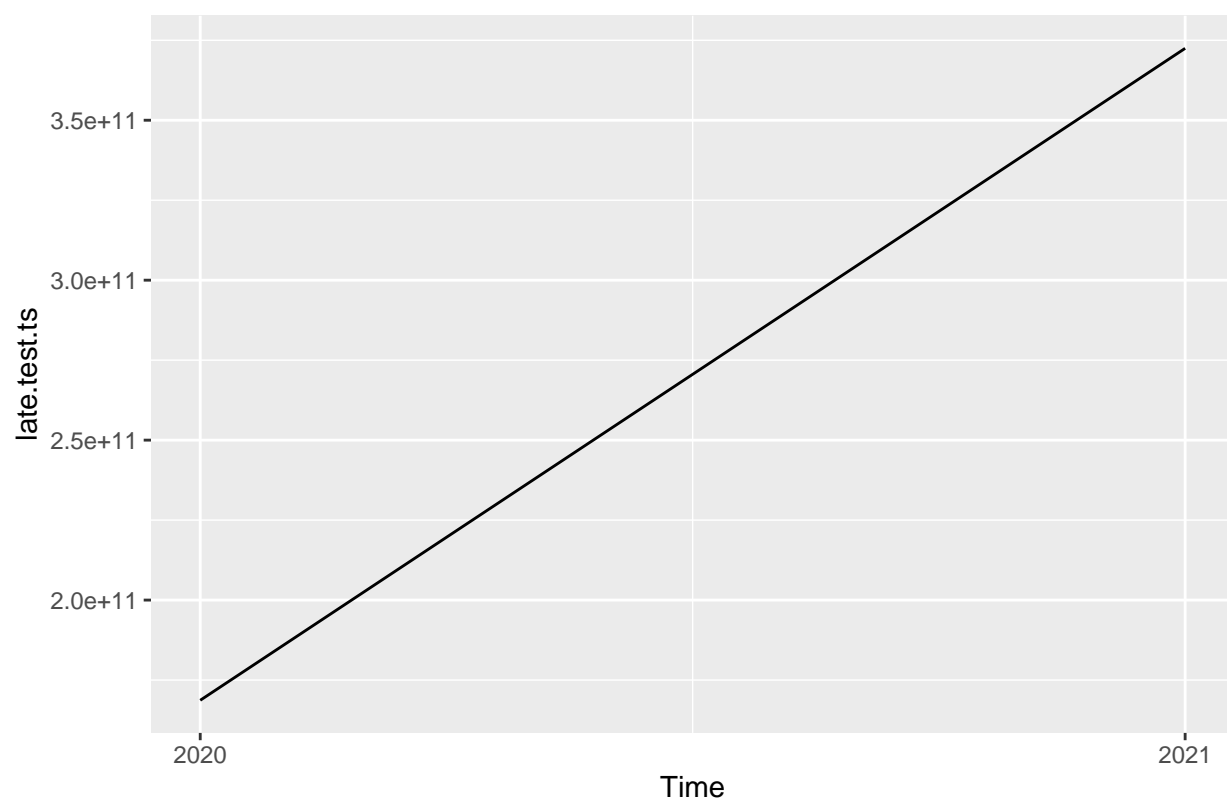
#Creating Time Series

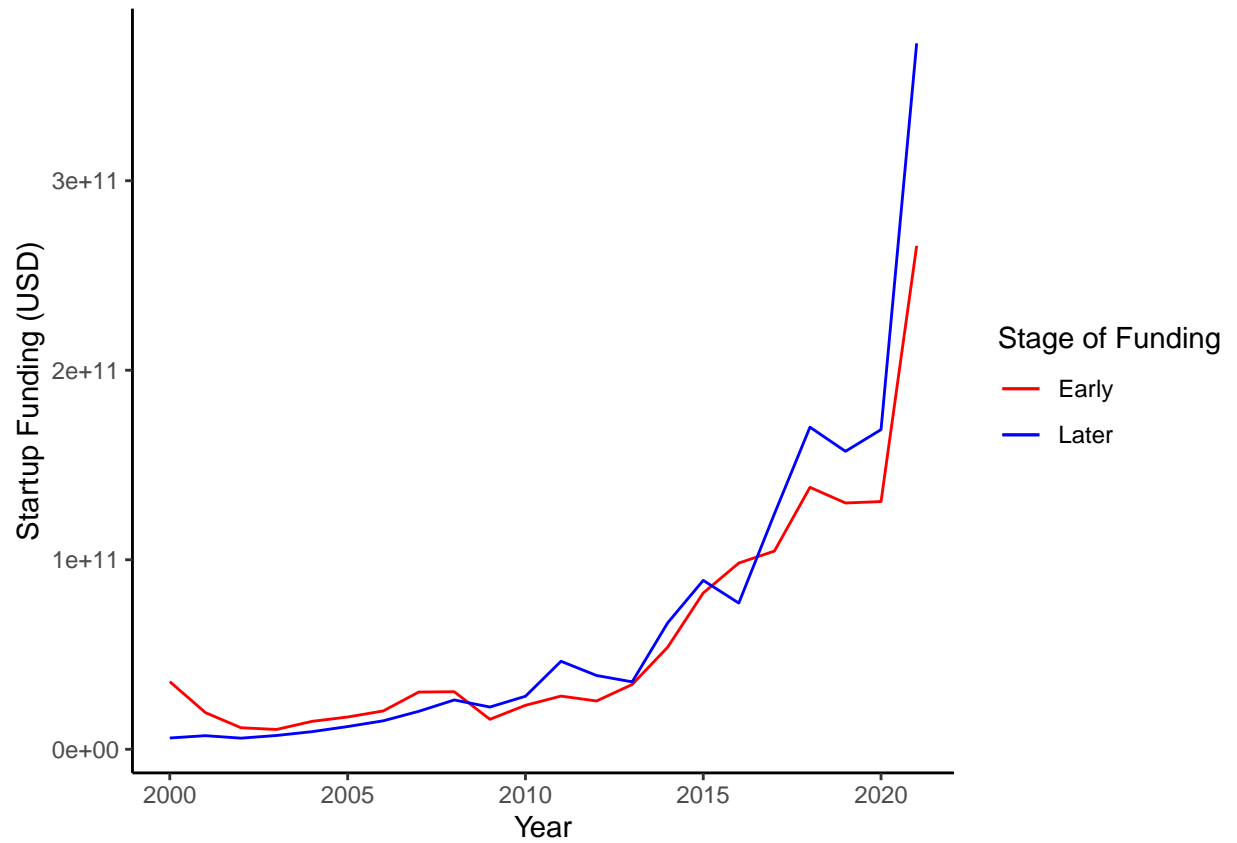
#Data Wrangling - Testing Sets The below graphs show the testing sets and the training sets for early and late stage.









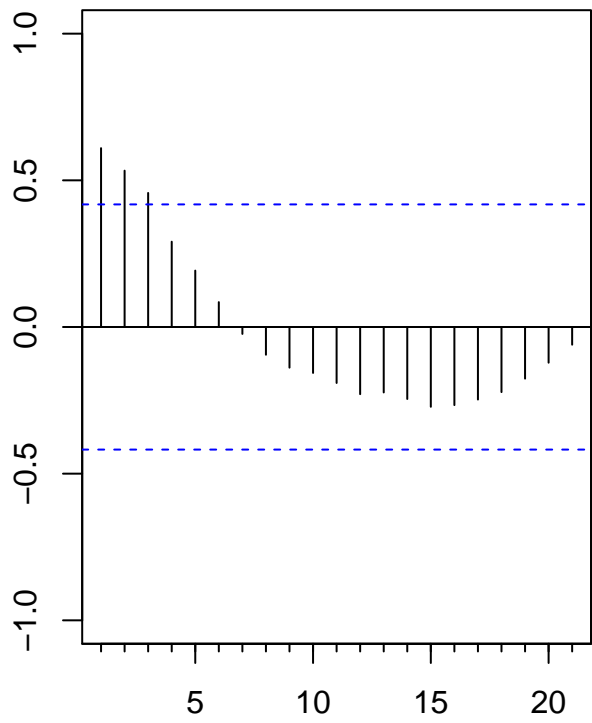


#Visualization

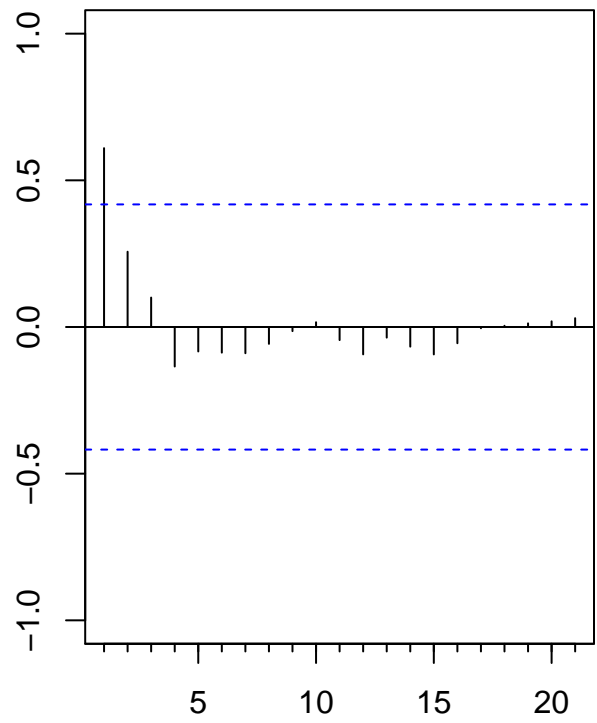
As you can see, both early and late stage funding has increased drastically over the years. Late stage passed early stage around 2016 and has been greater than early stage ever since.

#Create ACFs and PACFs

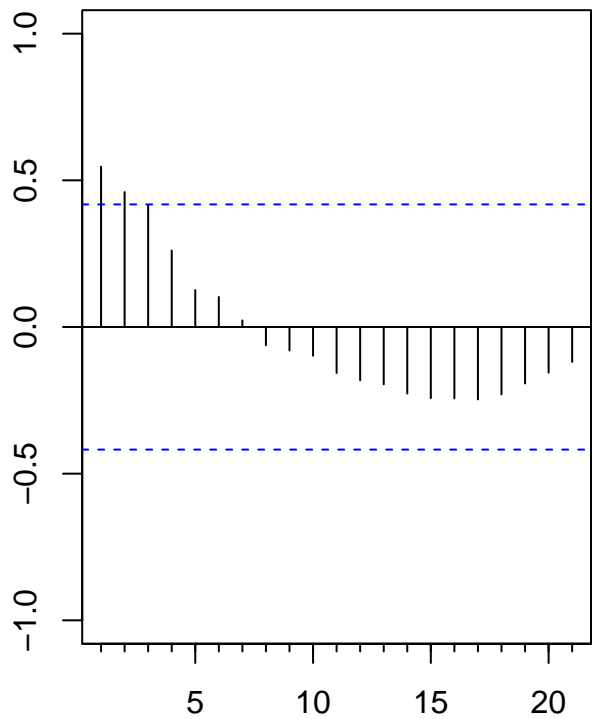
**Early ACF**



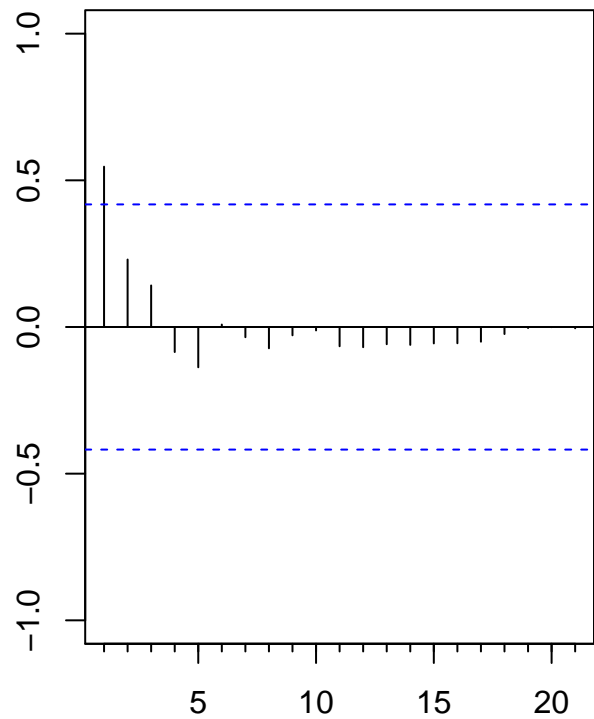
**Early PACF**



### Late ACF



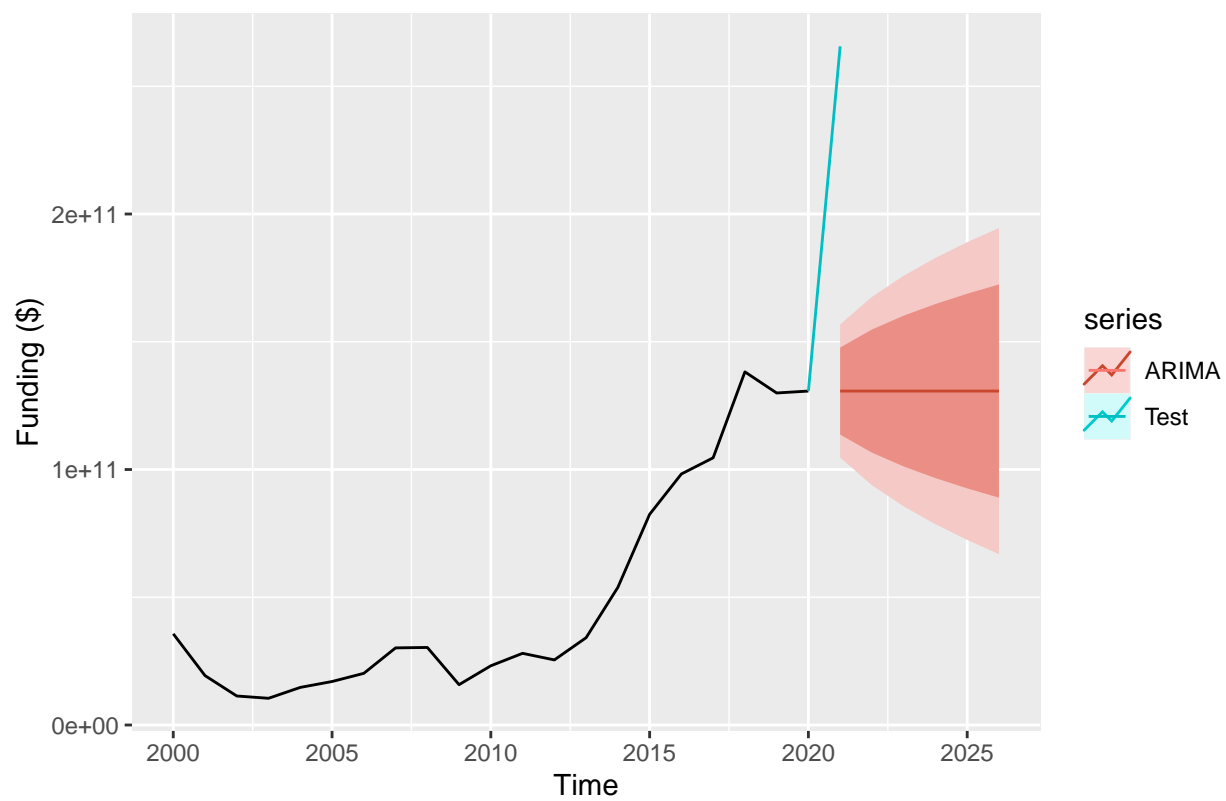
### Late PACF



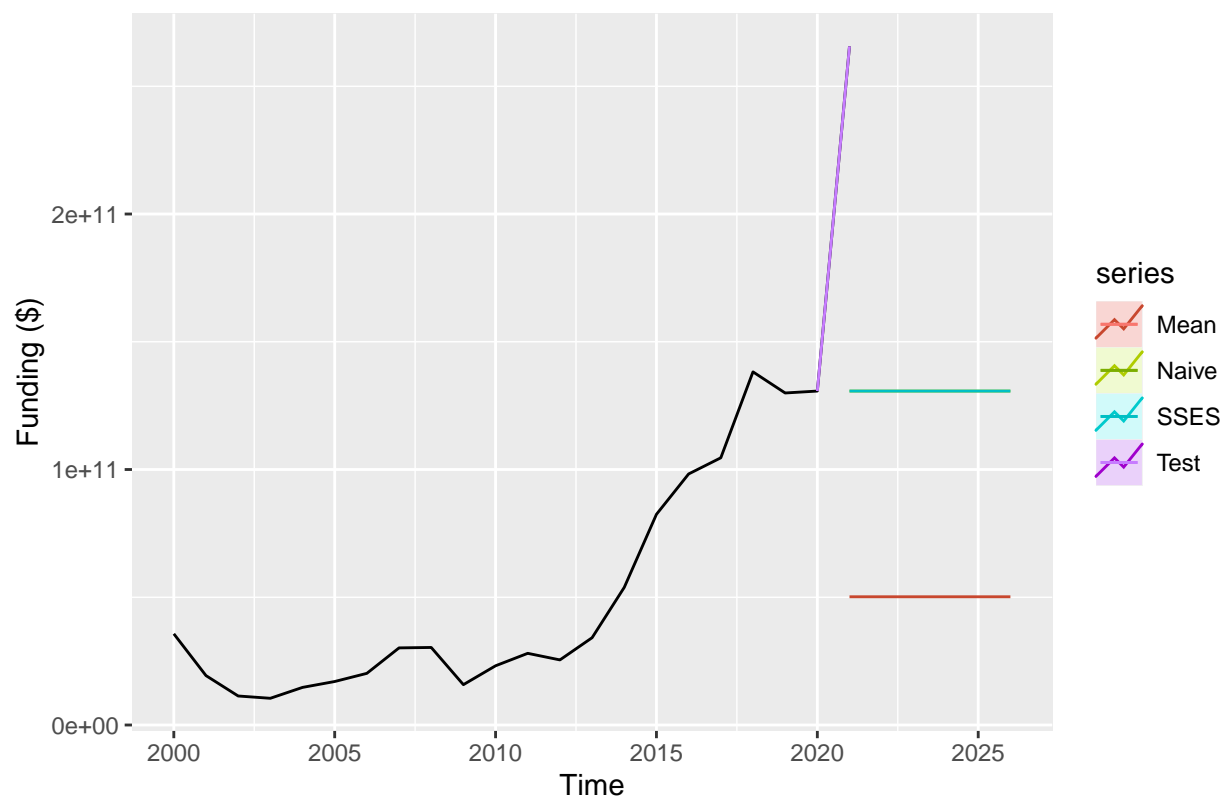
#Model Fit - Arima (Early)

```
## Series: early.training.ts
## ARIMA(0,1,0)
##
## sigma^2 = 1.767e+20: log likelihood = -494.59
## AIC=991.17 AICc=991.4 BIC=992.17
##
## Training set error measures:
##           ME          RMSE          MAE          MPE          MAPE          MASE
## Training set 4525751369 12971505571 9348854382 0.60679 26.52674 0.9525542
##           ACF1
## Training set 0.2714486
```



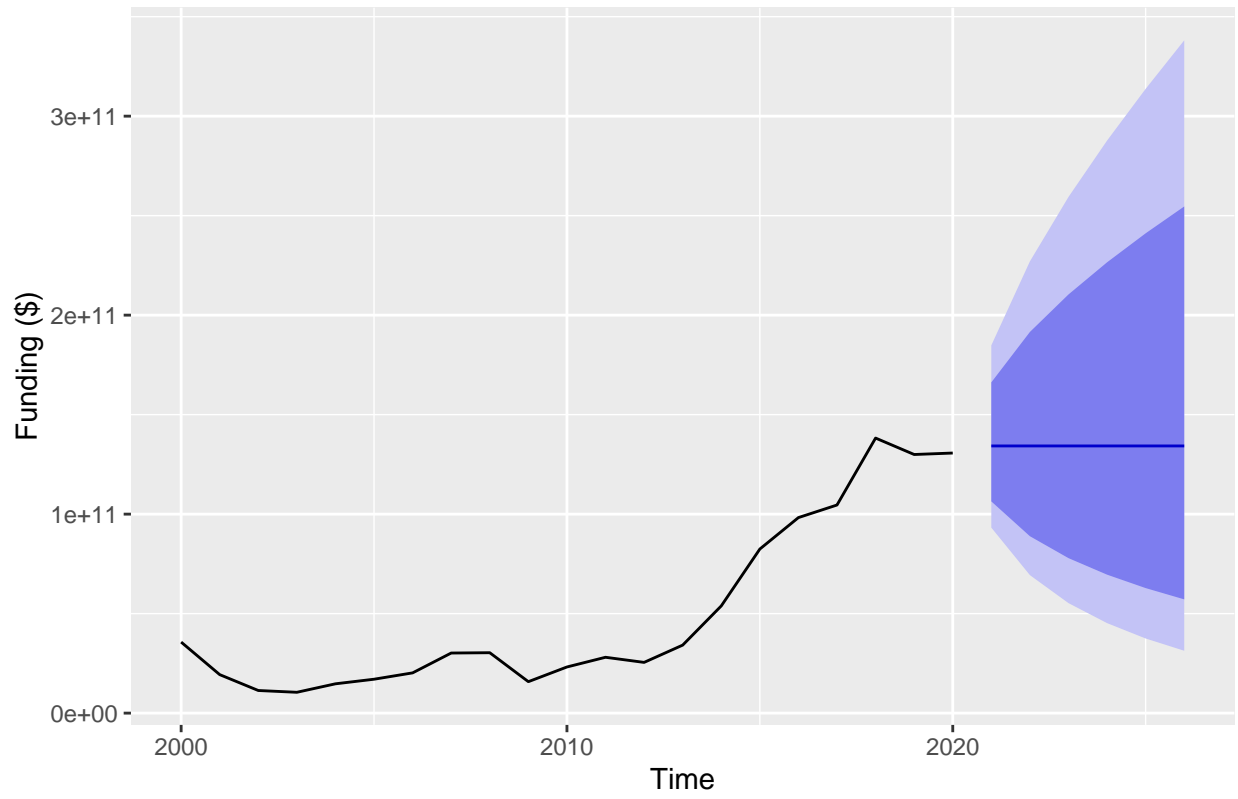


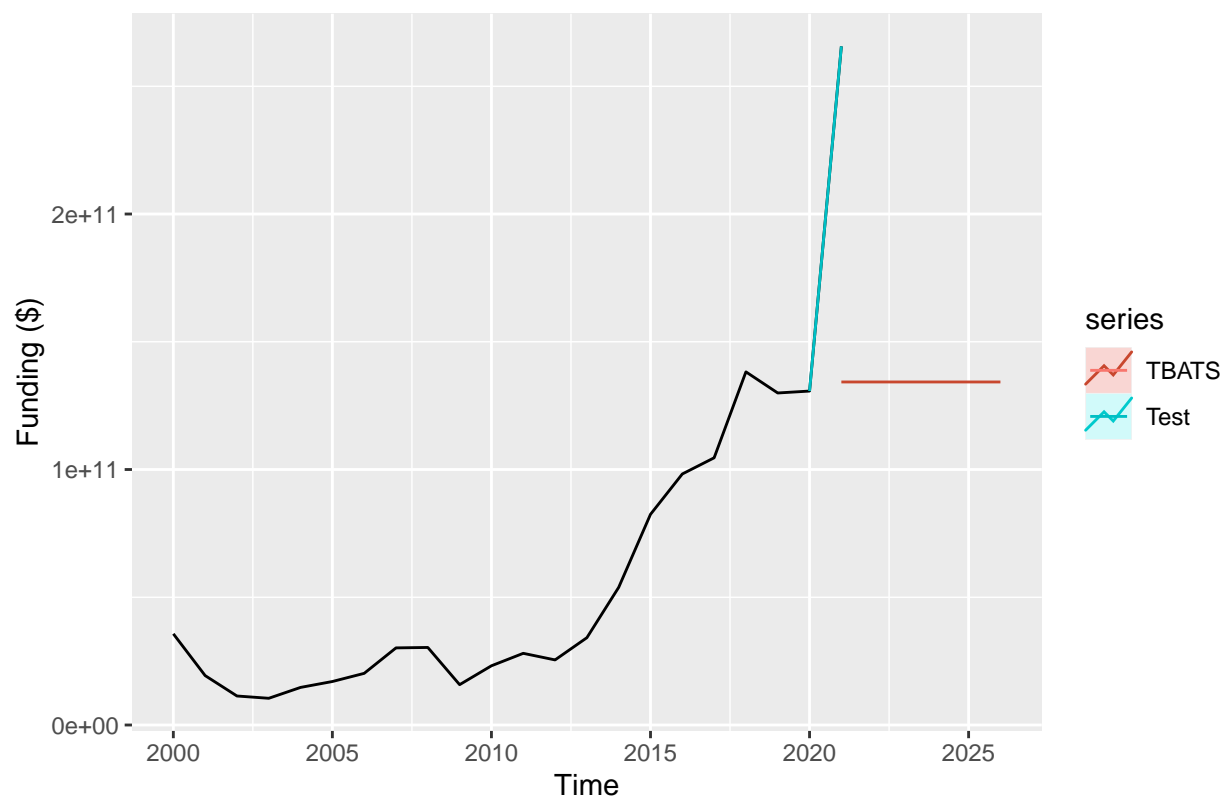
#Model Fit - Mean, Naive, SSES (Early)



#Model Fit - TBATS (Early)

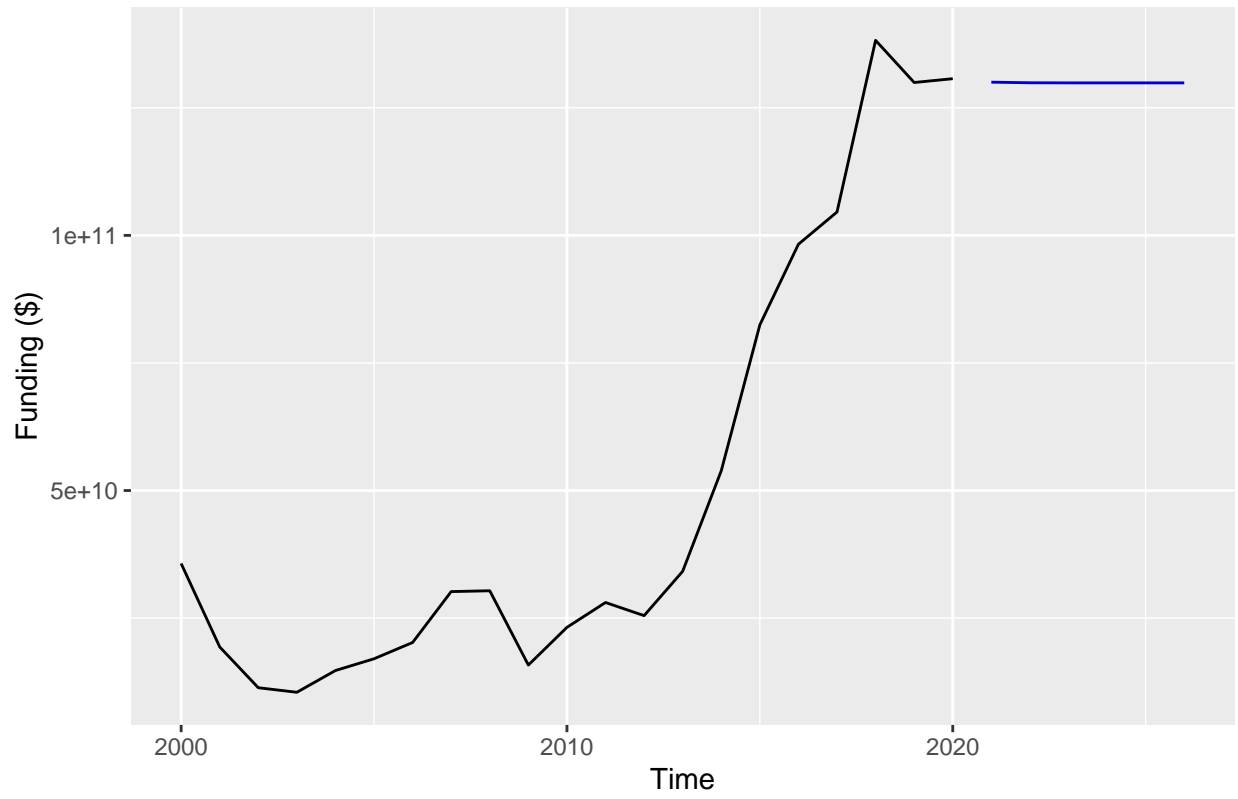
Forecasts from BATS(0.389, {0,0}, −, −)

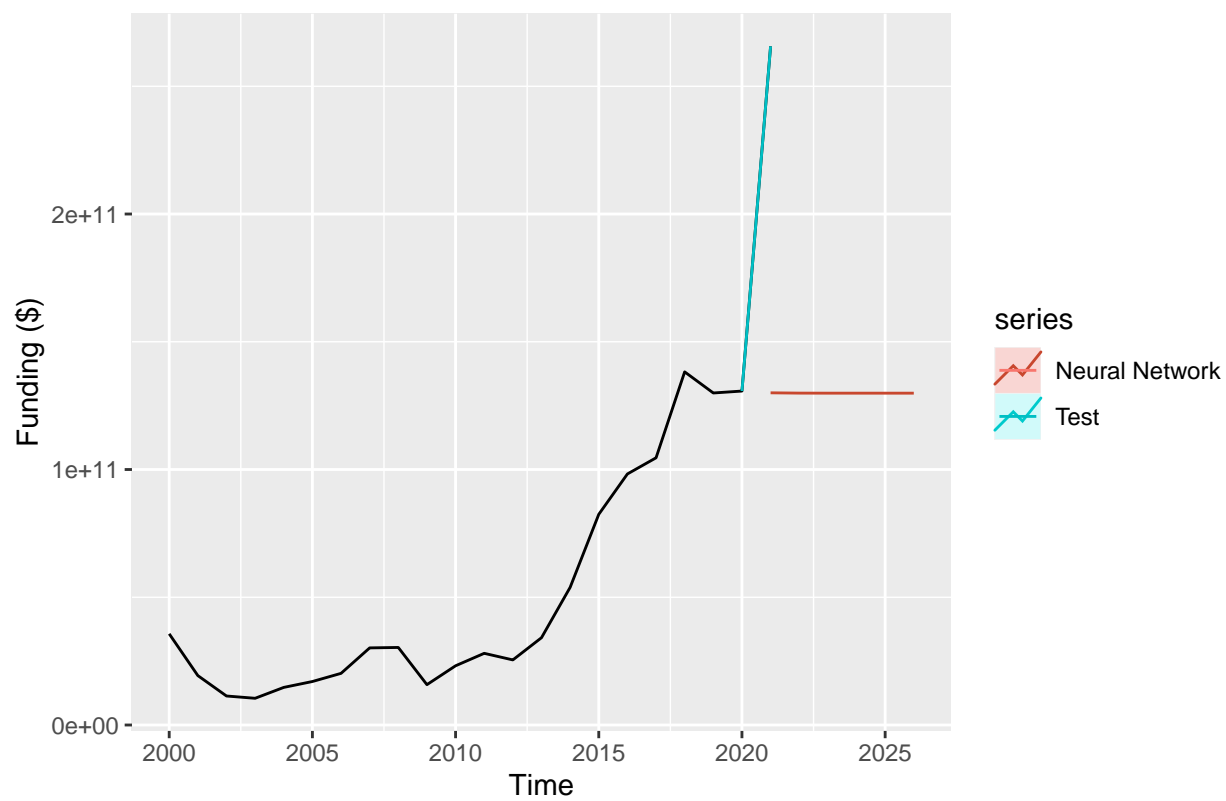




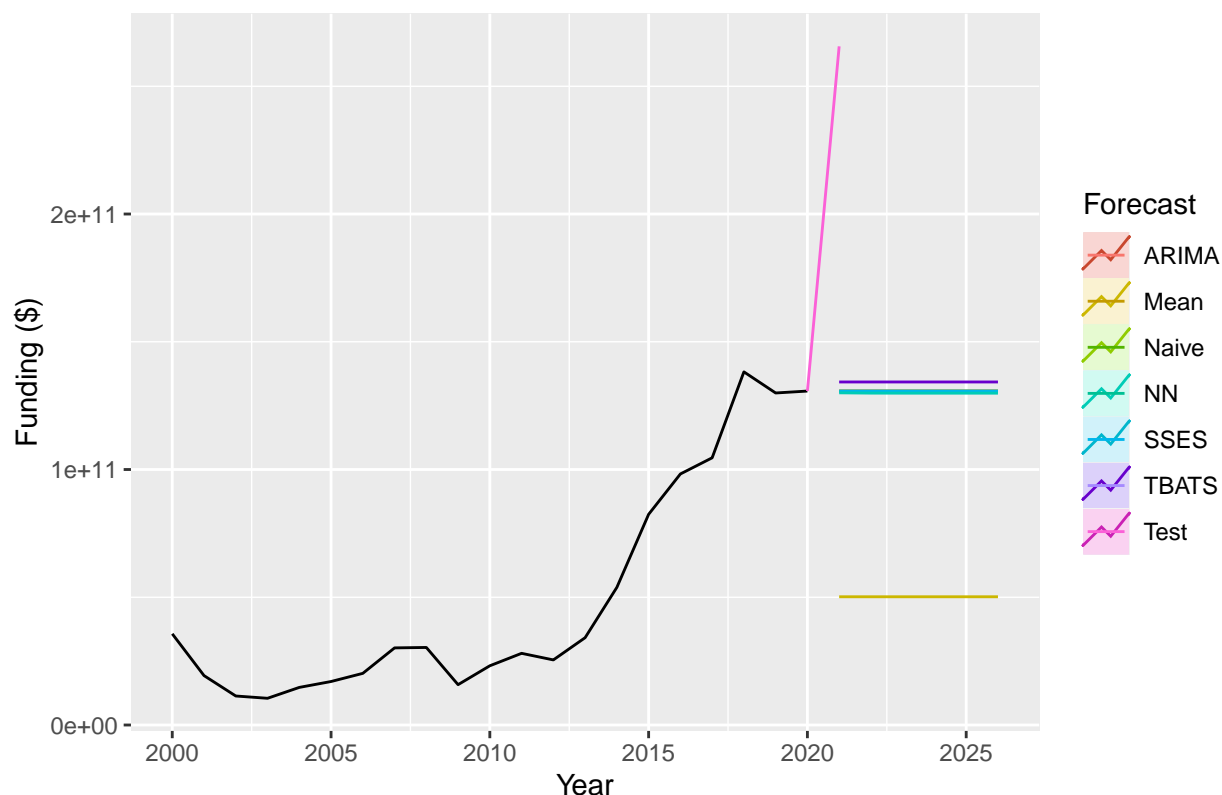
#Model Fit - NN (Early)

Forecasts from NNAR(1,1)





#Model Plotting (Early)



As you can tell from the graph, it is hard to see which model has the best fit for early stage investment. To better understand which model to use, we compared the RMSE scores. The model with the lowest RMSE score has the best fit.

#Model Scoring (Early)

	ME	RMSE	MAE	MPE	MAPE
ARIMA	134920037676	134920037676	134920037676	50.79304	50.79304
Mean	215430828331	215430828331	215430828331	81.10275	81.10275
Naive	134920037676	134920037676	134920037676	50.79304	50.79304
SSES	134920037676	134920037676	134920037676	50.79304	50.79304
TBATS	131357180623	131357180623	131357180623	49.45174	49.45174
Neural Network	135605264259	135605264259	135605264259	51.05100	51.05100

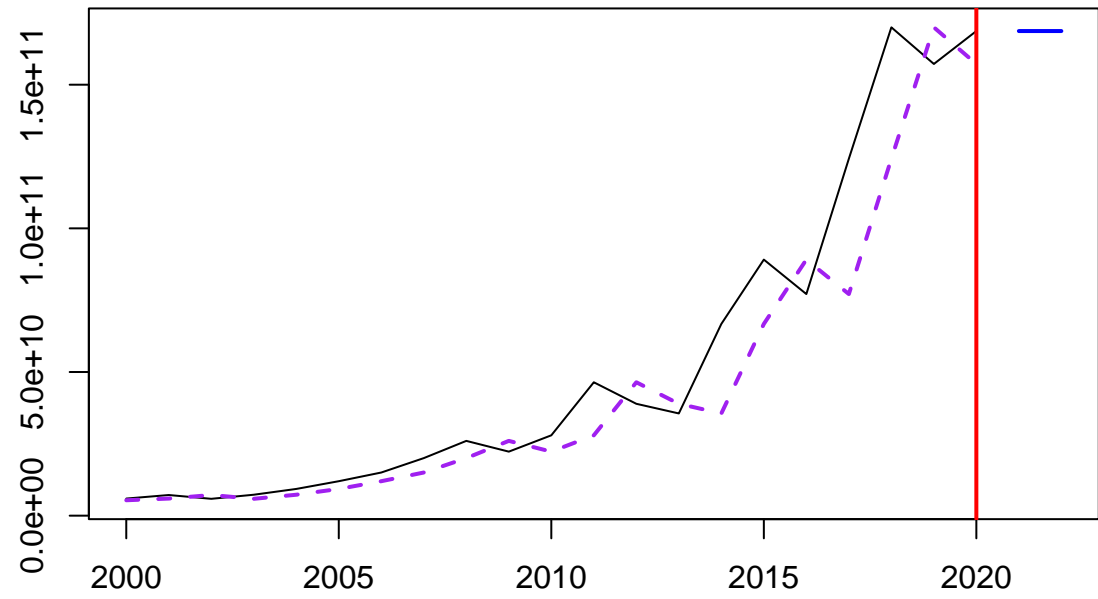
## The best model by RMSE is: TBATS

#RMSE - Early The below code is included in the kableExtra package. It runs in r markdown, but isn't compatible with knitting. We included it here to show our efforts.

```
#kbl(early.scores,
  #caption = "Forecast Accuracy for Early Stage Funding Date",
  #digits = array(6,ncol(early.scores))) %>%
  #kable_styling(full_width = FALSE, position = "center") %>%

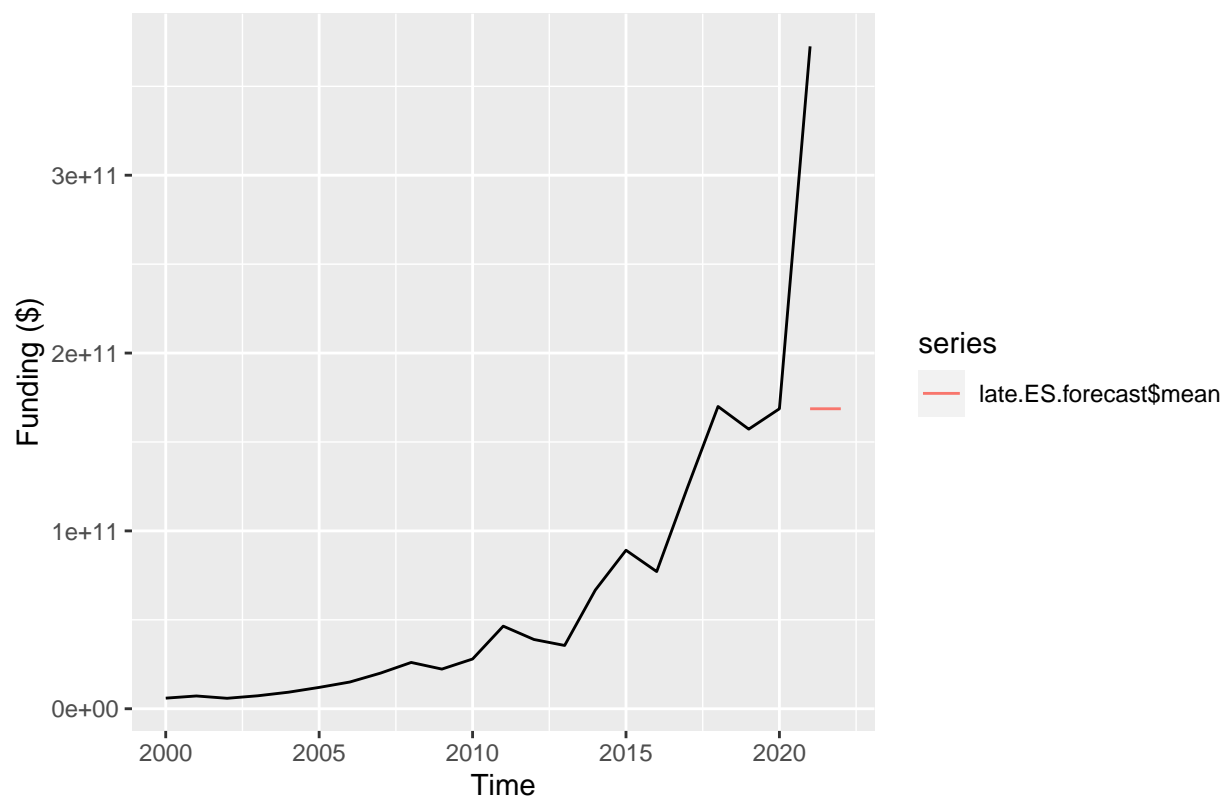
  #highlight model with lowest RMSE
  #kable_styling(latex_options="striped", stripe_index = which.min(early.scores[, "RMSE"]))
```

### Forecast from ETS(MNN) with Normal distribution



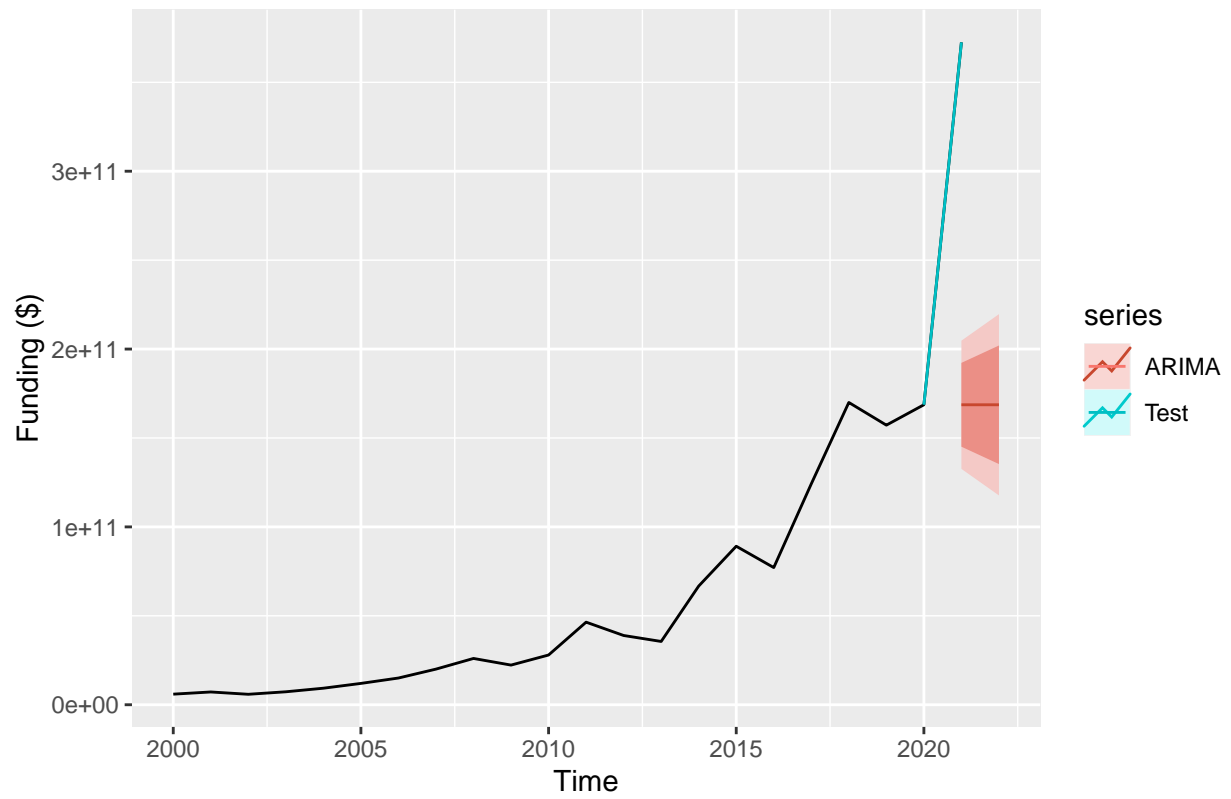
#Model Fit - ES (Late)



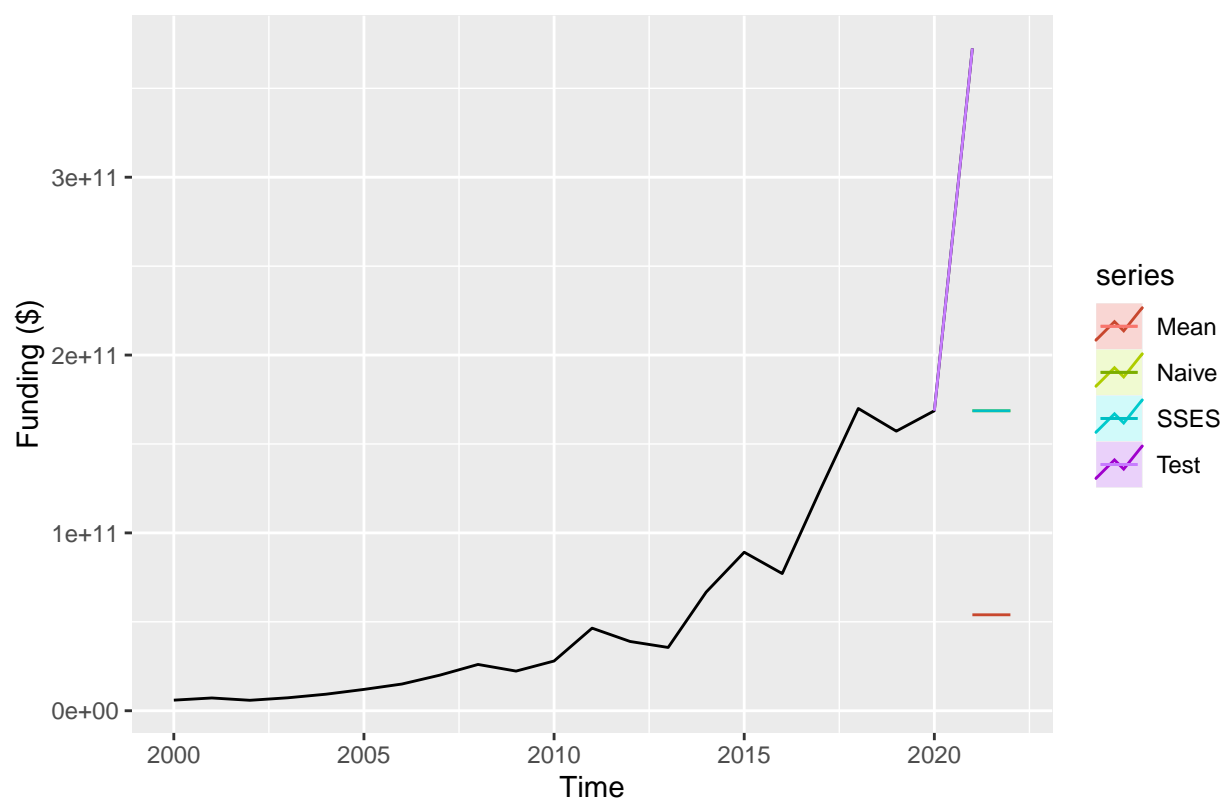


#Model Fit - Arima (Late)

```
## Series: late.training.ts
## ARIMA(0,1,0)
##
## sigma^2 = 3.38e+20: log likelihood = -501.08
## AIC=1004.15 AICc=1004.37 BIC=1005.15
##
## Training set error measures:
##           ME           RMSE           MAE           MPE           MAPE           MASE
## Training set 7748821973 17942595684 11619982072 12.42824 21.15022 0.9524042
##           ACF1
## Training set -0.007499492
```

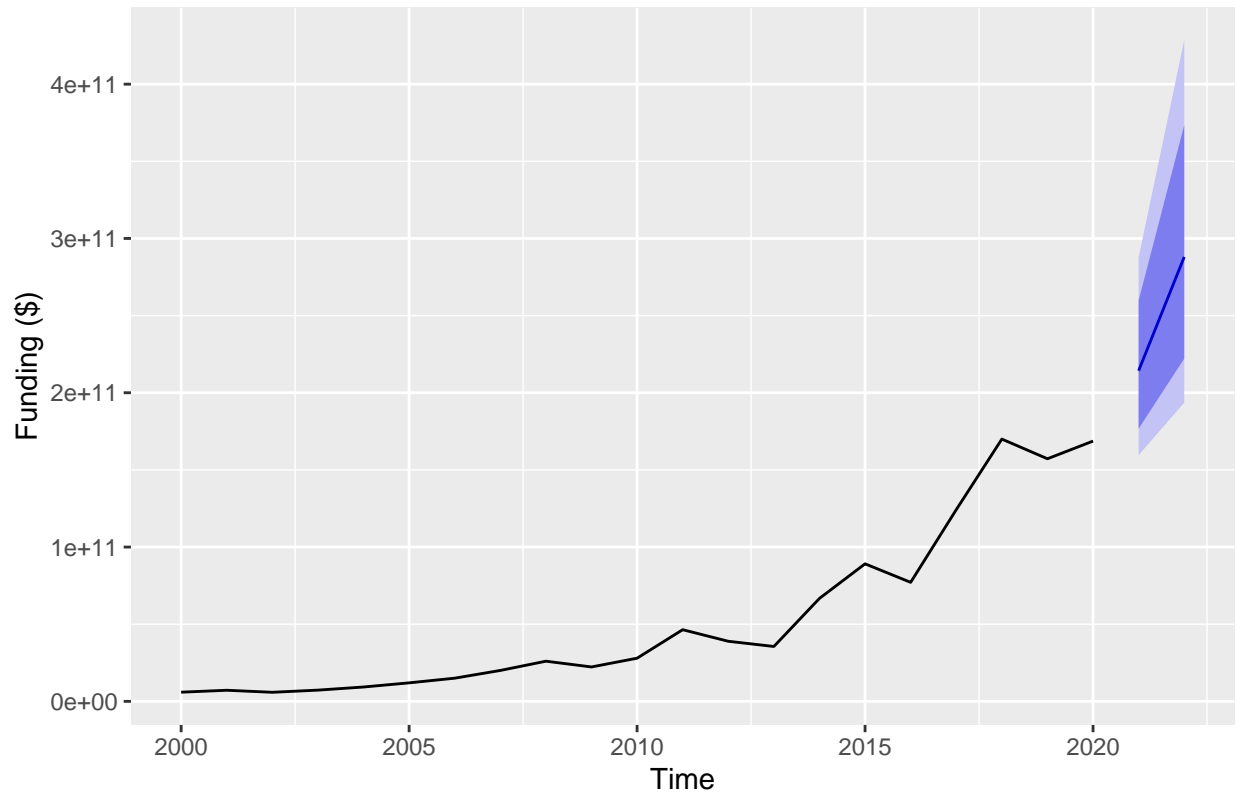


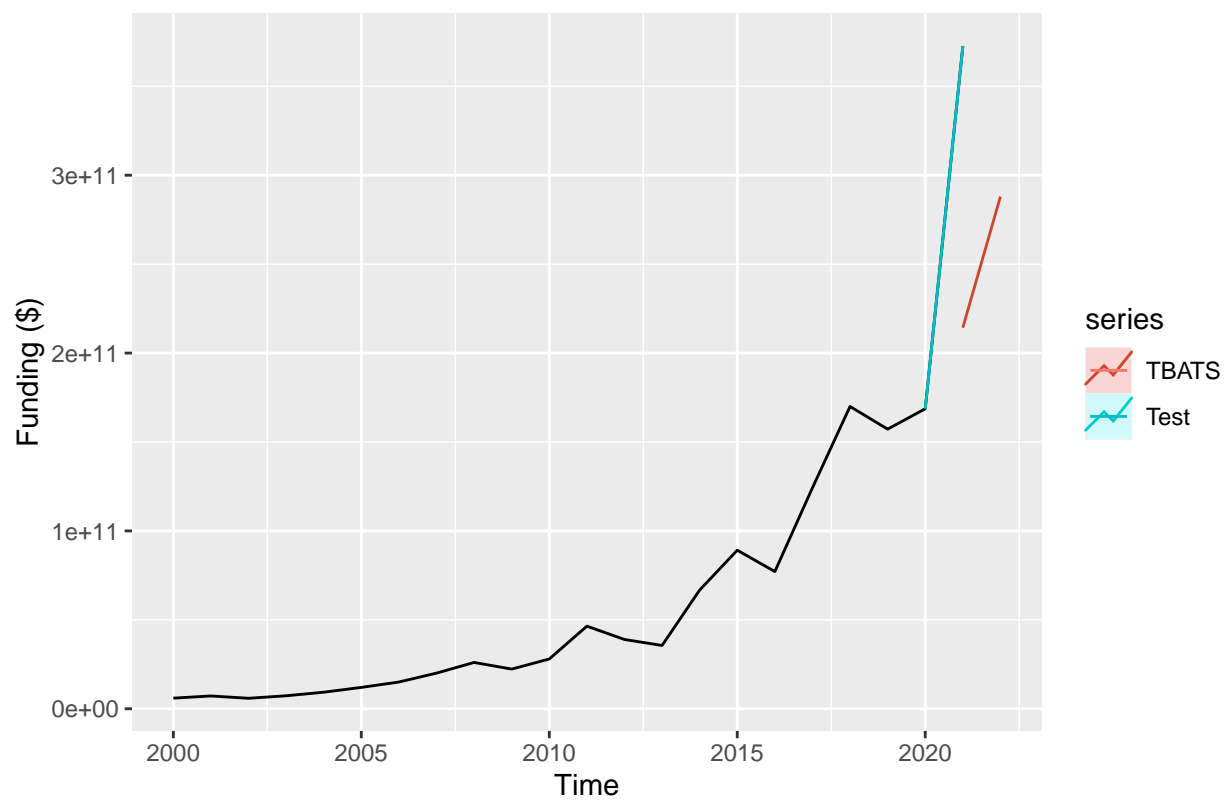
#Model Fit - Mean, Naive, SSES (Late)



#Model Fit - TBATS (Late)

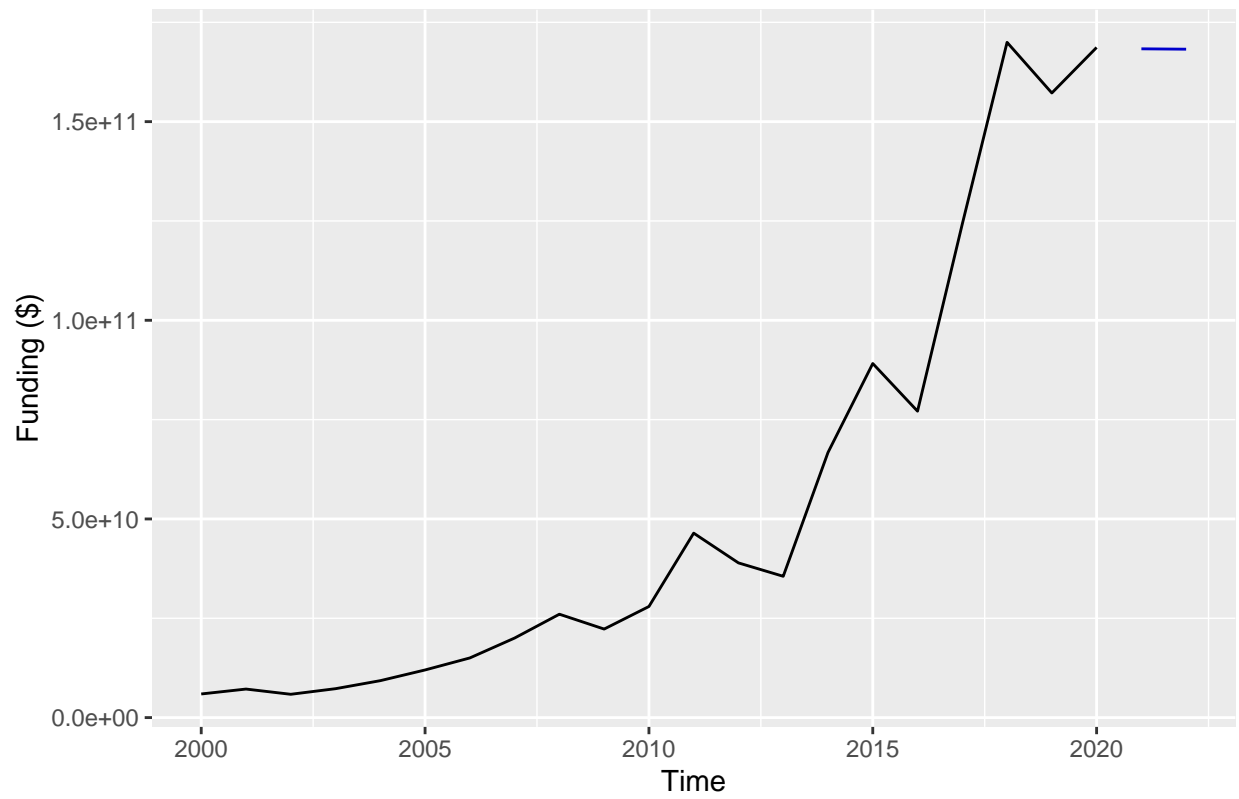
Forecasts from BATS(0.001, {0,1}, 1, -)

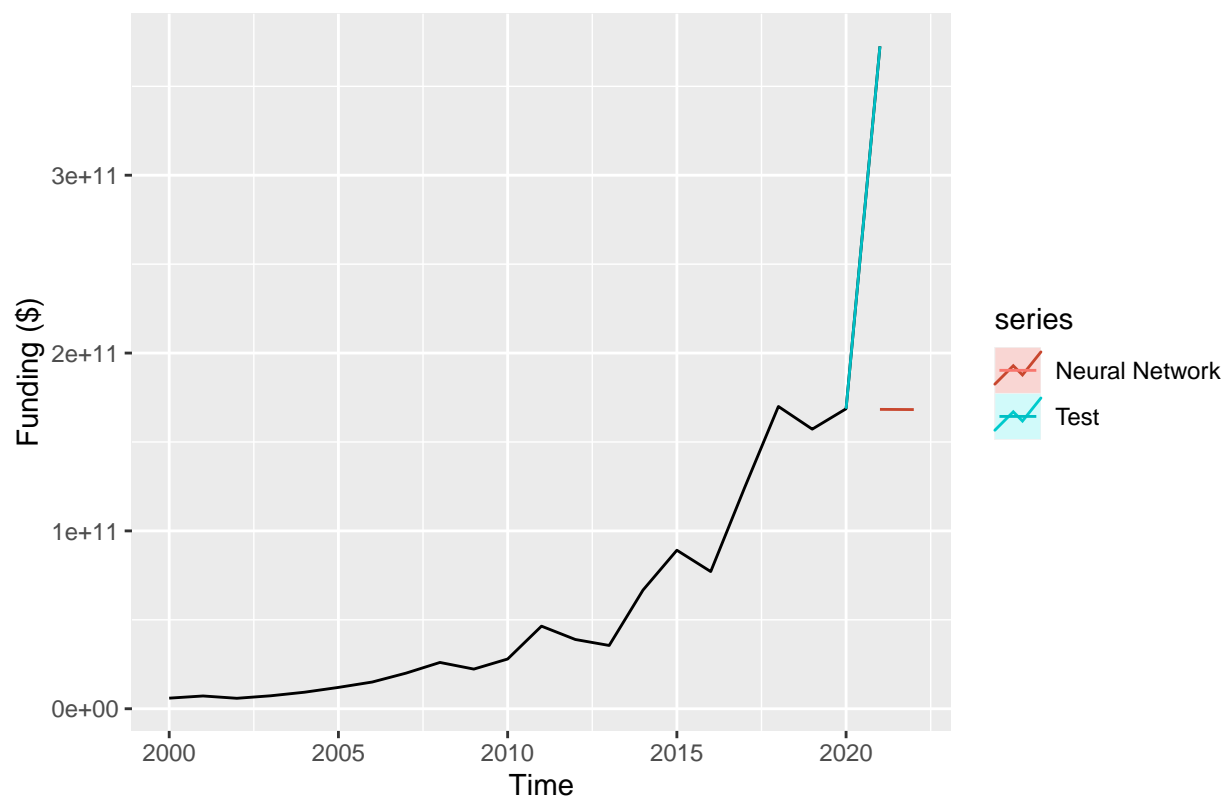




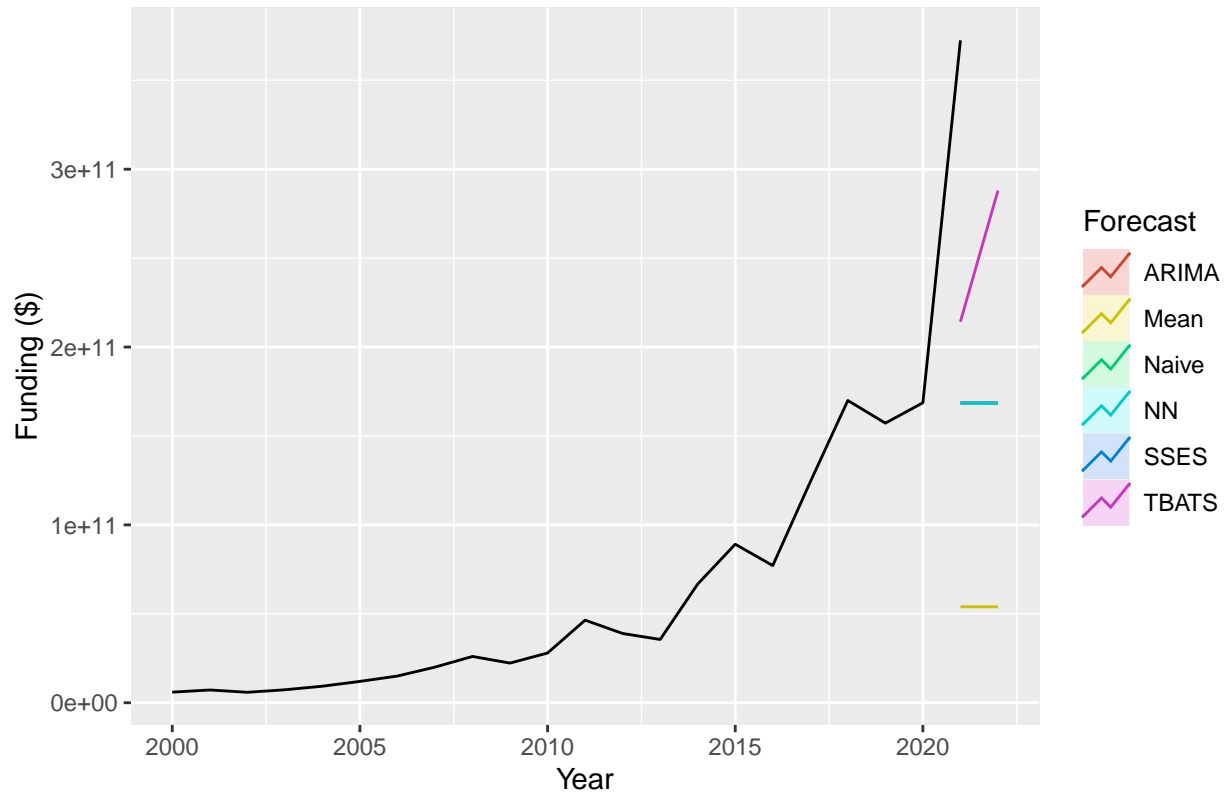
#Model Fit - NN (Late)

Forecasts from NNAR(1,1)





#Model Plotting (Late)



As you can tell from the graph, it is hard to see which model has the best fit for late stage investment. To better understand which model to use, we compared the RMSE scores. The model with the lowest RMSE score has the best fit.

#Model Scoring (Late)

	ME	RMSE	MAE	MPE	MAPE
ARIMA	203812130292	203812130292	203812130292	54.71748	54.71748
Mean	318535727367	318535727367	318535727367	85.51734	85.51734
Naive	203812130292	203812130292	203812130292	54.71748	54.71748
SSES	203812130292	203812130292	203812130292	54.71748	54.71748
TBATS	158210031971	158210031971	158210031971	42.47468	42.47468
Neural Network	204154534260	204154534260	204154534260	54.80940	54.80940

## The best model by RMSE is: TBATS

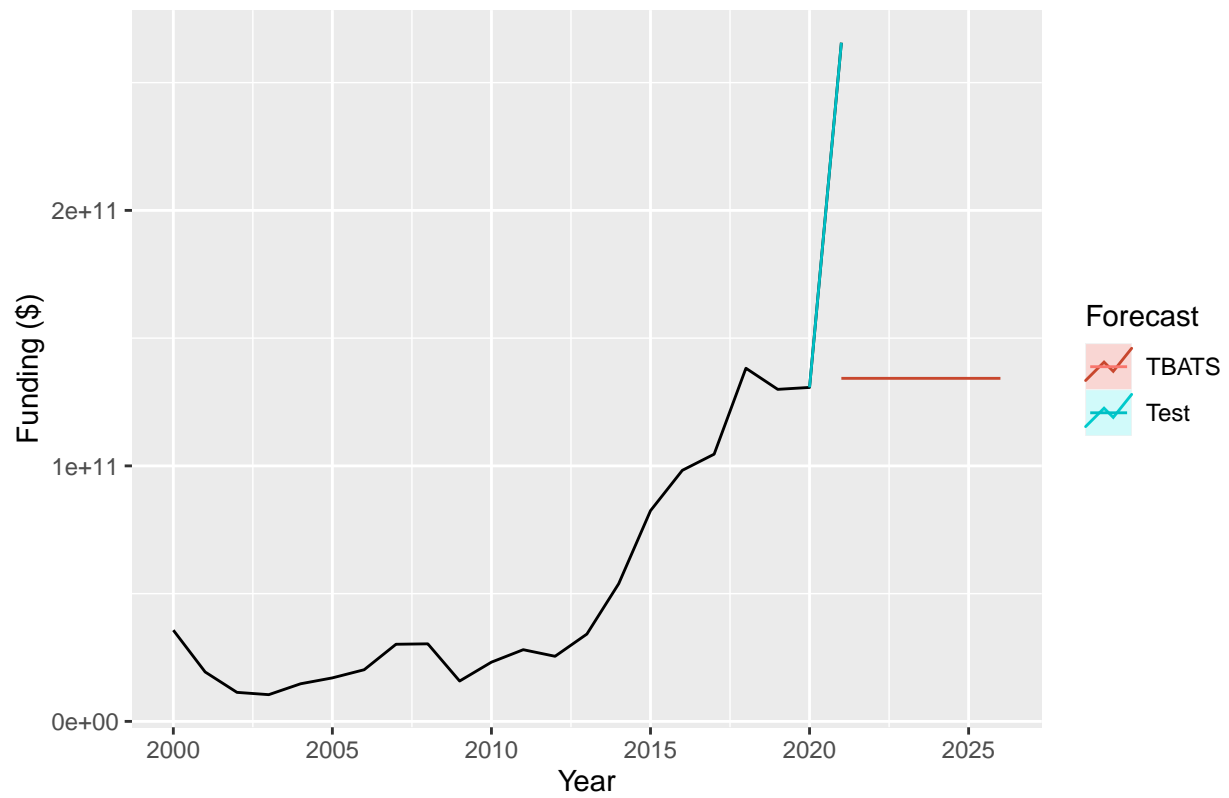
#RMSE - Late The below code is included in the kableExtra package. It runs in r markdown, but isn't compatible with knitting. We included it here to show our efforts.

```
#kbl(late.scores,
  #caption = "Forecast Accuracy for Late Stage Funding Date",
  #digits = array(6,ncol(late.scores))) %>%
  #kable_styling(full_width = FALSE, position = "center") %>%

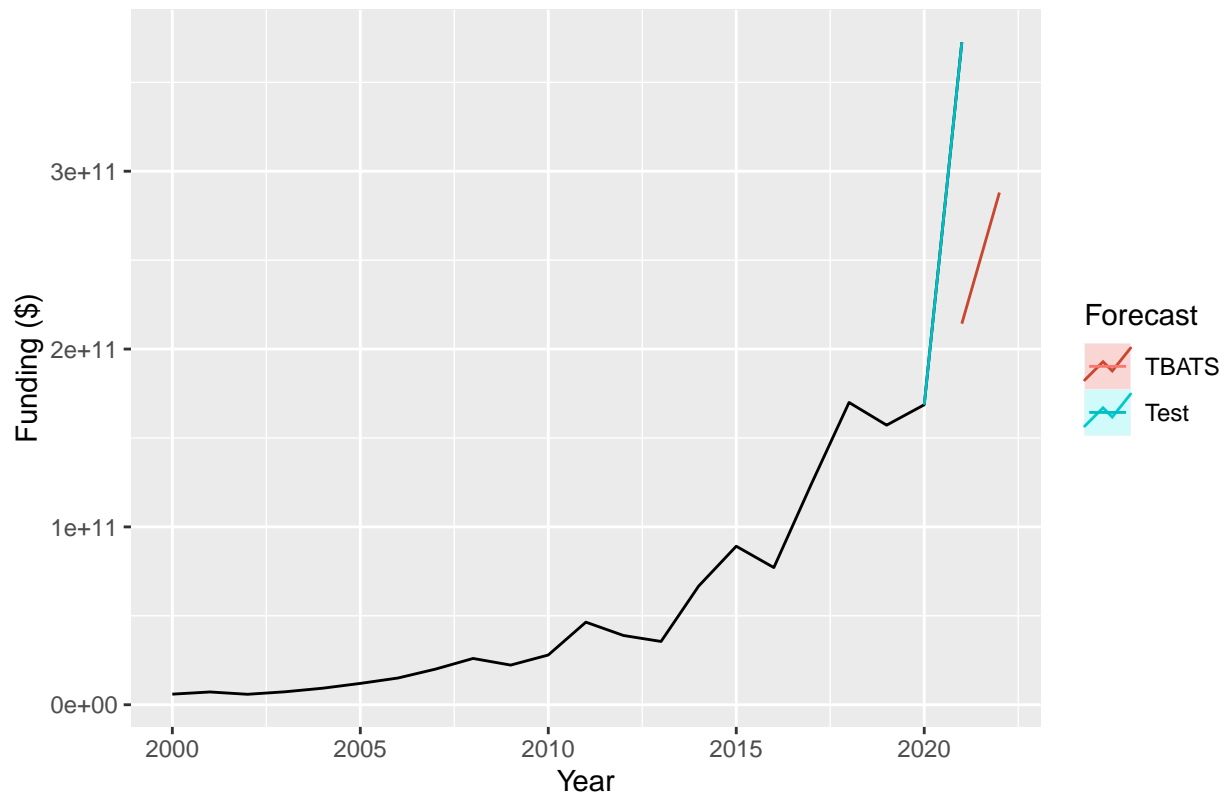
  #highlight model with lowest RMSE
  #kable_styling(latex_options="striped", stripe_index = which.min(late.scores[, "RMSE"]))
```



#Plot the Best Model (Early)



#Plot the Best Model (Late)



**Conclusions** There are some limitations to this dataset which have made it difficult to predict. These limitations made predictions difficult. It is unclear whether there will continue to be an increase in clean tech startup investment in the future. Looking at the “best” model, the forecast for early stage investment seems to stay constant (a straight line), where late stage investment has an increase. This shows that, with the limited data available, there will be an overall increase in clean tech investment and the late stage investment will continue to be greater than early stage investment.

**Possible Future Steps** Analyze the breakdown of early vs. late-stage investment by sector (energy, non-energy, grid, fossil fuel, energy efficiency, etc.)