Netflix Subscribers Forecasting

Anna Atkuru

Introduction

- Netflix had just 21.4 million paid subscribers in 2011. In 7 years, that number is **more than 7x** as high. The data tells us a story of how streaming platforms came to popularity over time.
- Now, at approximately **158.33 million paid subscribers worldwide**, Netflix faces rivals such as Hulu, Disney+, Amazon and other streaming platforms that are growing at a rapid rate.
- Even with growing rivals, Netflix continues to see an increase in subscribers over time. This shows how popular streaming platforms as a whole are becoming, and how Netflix's strategies for growth continue to be successful.

Objectives

- With the rise of rivals, Netflix must be able analyze its growth thus far, and be able to forecast future performance.
- Using historical data of **additional paid subscribers per quarter**, we used time series techniques to forecast Netflix's growth.
- Factors such as **price hikes**, **rival growth** and **popular show releases** have affected Netflix's overall performance.
- Allowing past trends to predict future growth will allow for consistent subscriber growth.

Data Description

- There are **35** total data points in our dataset
  - For the purpose of forecasting, our **training data** comprised of **additional paid subscribers** from Q1 2011 - Q4 2017.
  - Our **testing dataset** comprised of additional paid subscribers from Q1 2018 - Q3 2019
- Each fiscal year is separated in 4 quarters.
  - Ranges from Q1 2011 to Q3 2019

MEAN	MEDIAN	STD. DEV.	VARIANCE
3.952	3.83	2.483653381	6.168534118

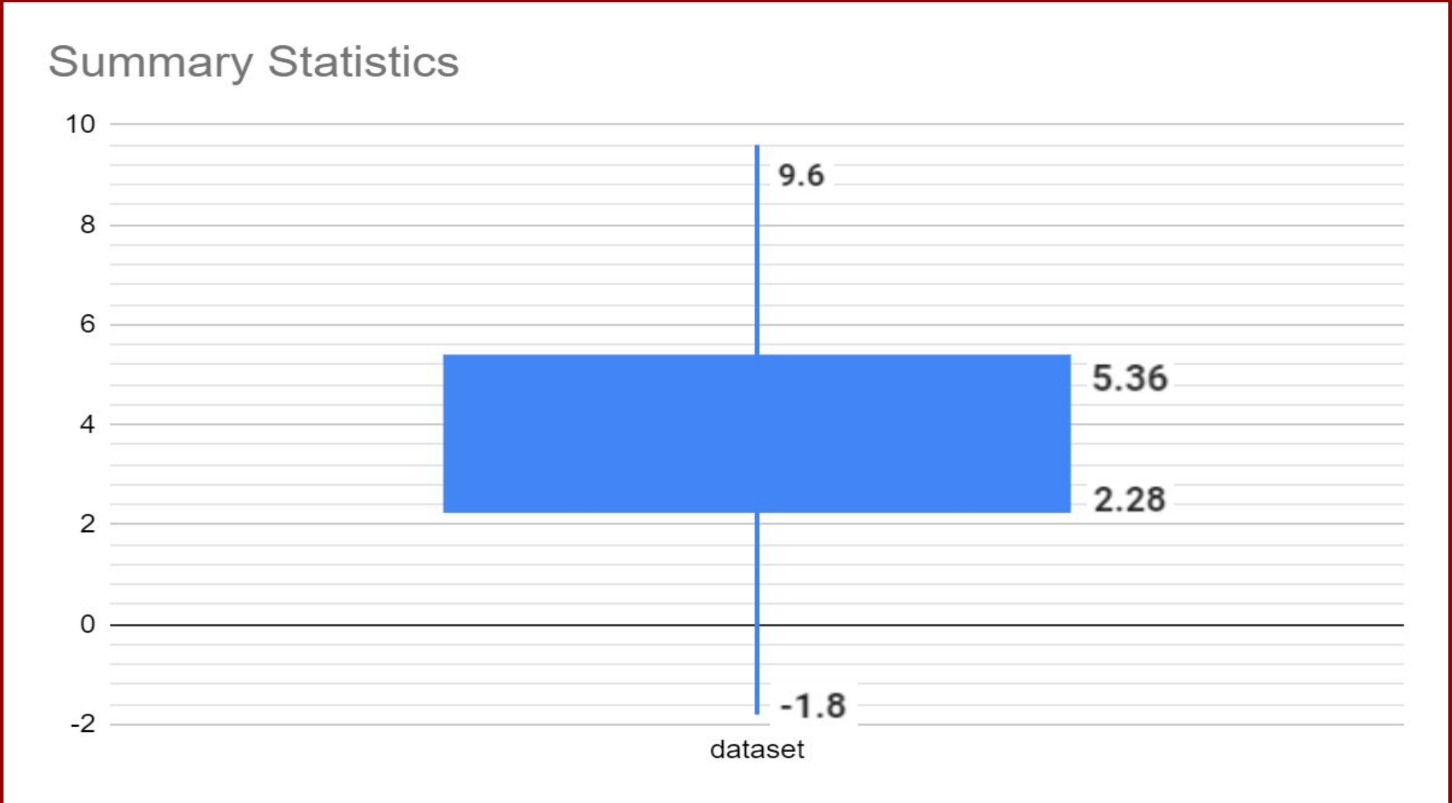
```
> summary(subscribers)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -1.800   2.280   3.830   3.952   5.360   9.600
```

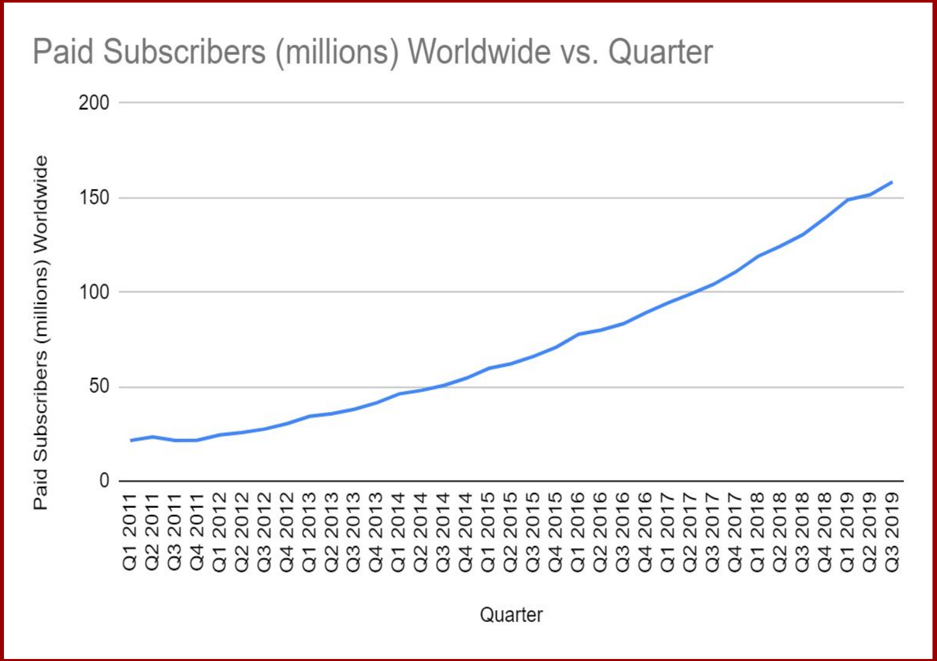
Netflix Additional Paid Subscribers (2011 - 2019)

Data Set	Quarter 1	Quarter 2	Quarter 3	Quarter 4
2011	1.39	1.9	-1.8	0.1
2012	2.83	1.28	1.78	2.87
2013	3.88	1.4	2.37	3.42
2014	4.7	1.86	2.66	3.83
2015	5.14	2.46	3.94	4.82
2016	6.87	2.19	3.38	5.81
2017	5.27	4.68	4.98	6.62
2018	8.26	5.45	6.07	8.84
2019	9.6	2.7	6.77	

**reported in millions*

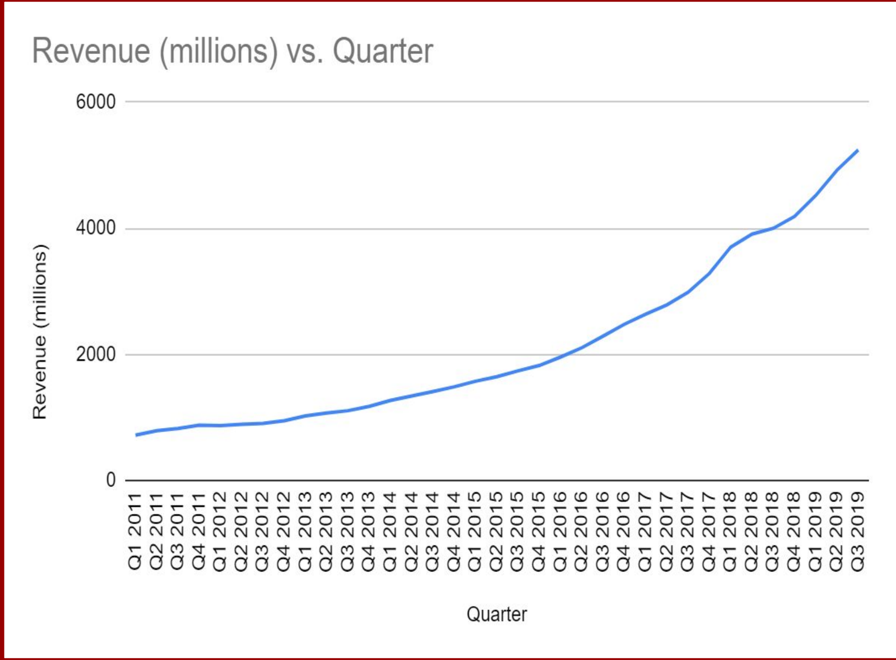
Boxplot of Change in Subscribers

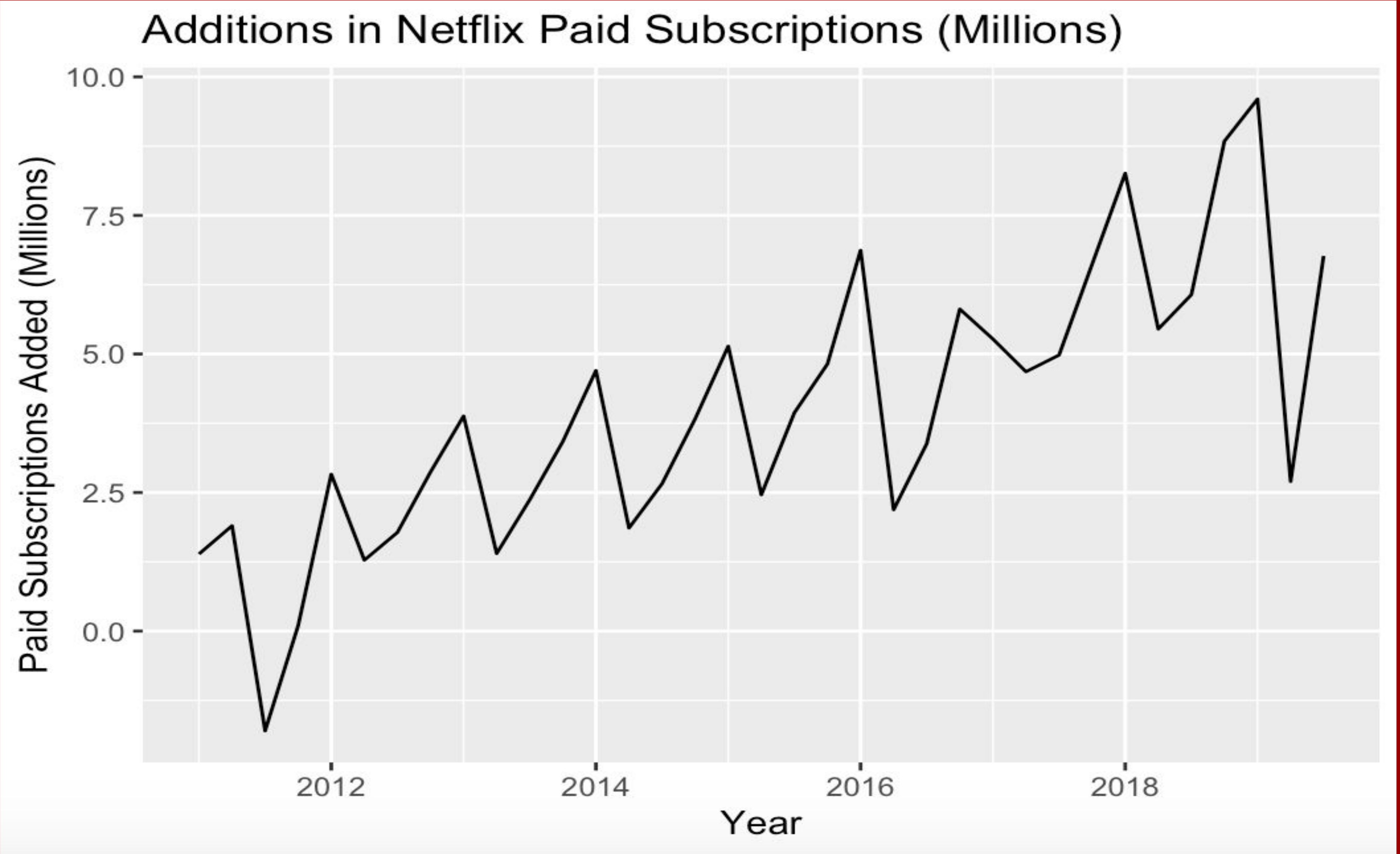




As subscribers increase over time...

...so does Netflix's revenue per quarter

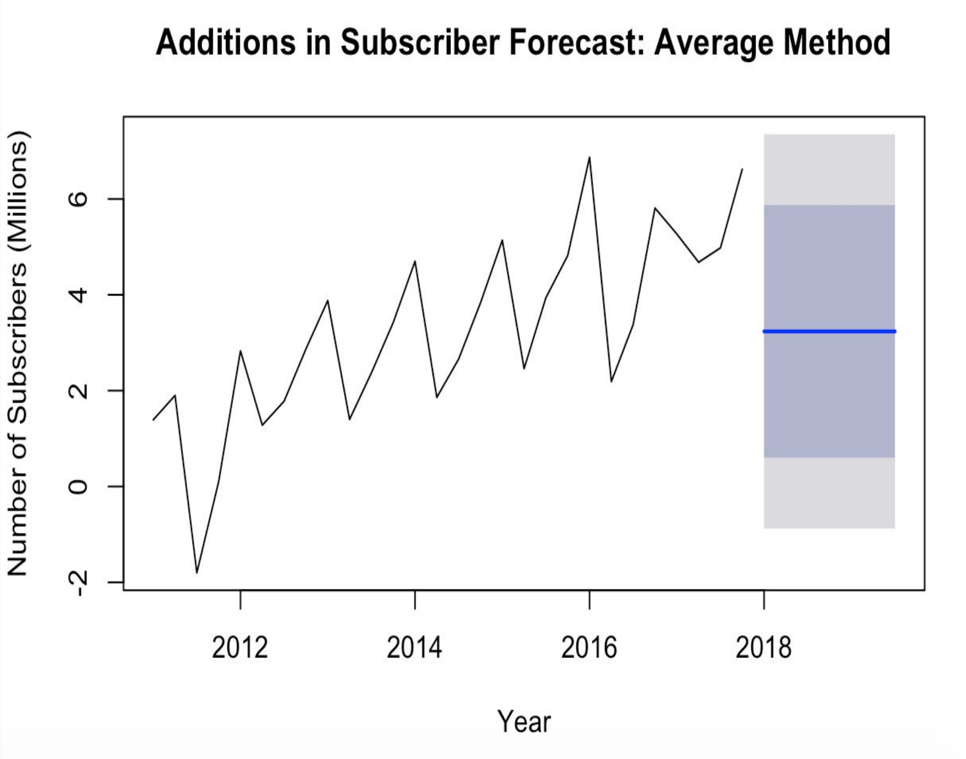



Additions in Netflix Paid Subscriptions (Millions)
Paid Subscriptions Added (Millions)
10.0
7.5
5.0
2.5
0.0
2012
2014
2016
2018
Year

Time Series Techniques

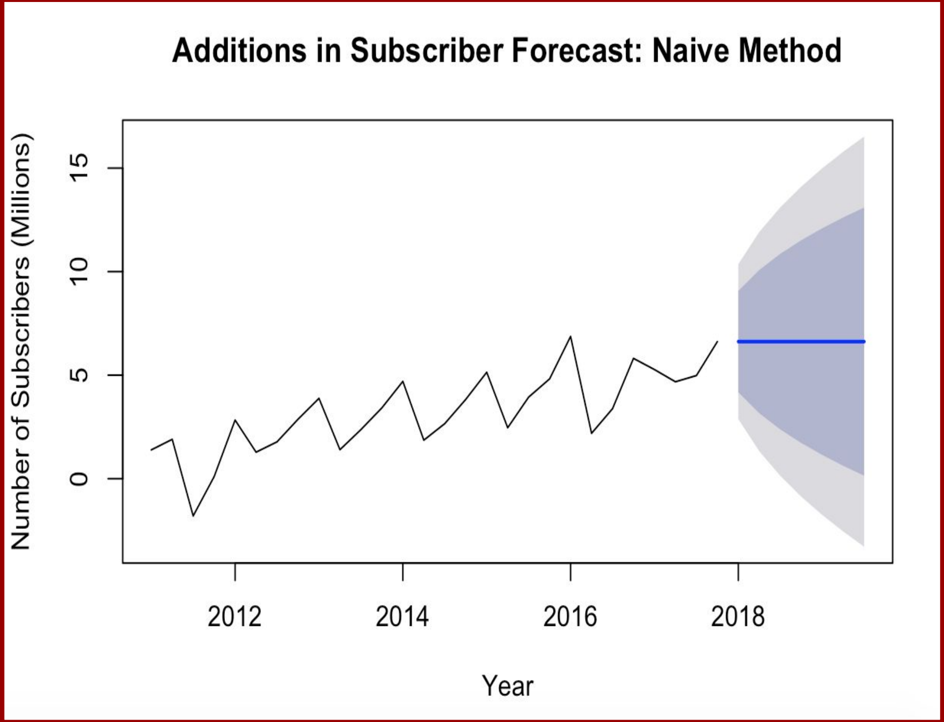
Forecasting: Average Method

```
> #SIMPLE AVG
> simpleaverage = meanf(subscribers,7)
> simpleaverage
        Point Forecast     Lo 80    Hi 80      Lo 95    Hi 95
2018 Q1       3.236786 0.6025661 5.871005 -0.8775175 7.351089
2018 Q2       3.236786 0.6025661 5.871005 -0.8775175 7.351089
2018 Q3       3.236786 0.6025661 5.871005 -0.8775175 7.351089
2018 Q4       3.236786 0.6025661 5.871005 -0.8775175 7.351089
2019 Q1       3.236786 0.6025661 5.871005 -0.8775175 7.351089
2019 Q2       3.236786 0.6025661 5.871005 -0.8775175 7.351089
2019 Q3       3.236786 0.6025661 5.871005 -0.8775175 7.351089
```



Forecasting: Naive Method

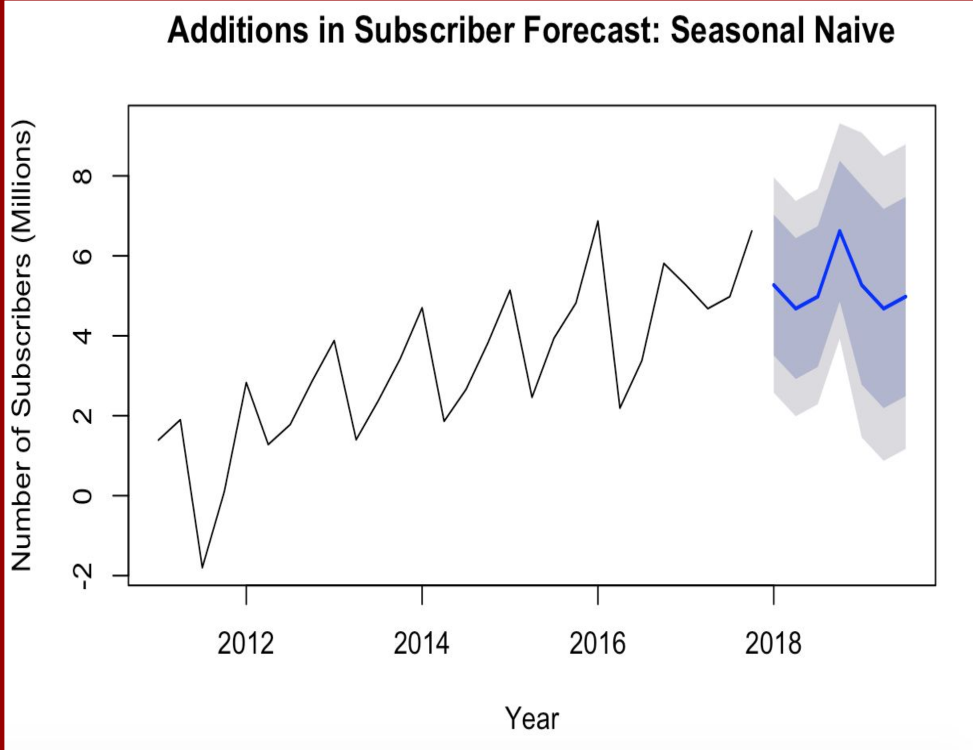
```
> #NAIVE
> naivemethod = naive(subscribers,7)
> naivemethod
        Point Forecast     Lo 80     Hi 80      Lo 95    Hi 95
2018 Q1           6.62 4.1723583  9.067642  2.8766550 10.36334
2018 Q2           6.62 3.1585119 10.081488  1.3261108 11.91389
2018 Q3           6.62 2.3805602 10.859440  0.1363363 13.10366
2018 Q4           6.62 1.7247166 11.515283 -0.8666900 14.10669
2019 Q1           6.62 1.1469067 12.093093 -1.7503738 14.99037
2019 Q2           6.62 0.6245267 12.615473 -2.5492851 15.78929
2019 Q3           6.62 0.1441487 13.095851 -3.2839599 16.52396
```



Forecasting: Seasonal Naive Method

```
> #SEASONAL NAIVE
> seasonalnaivemethod = snaive(subscribers, 7, freq=4)
> seasonalnaivemethod
```

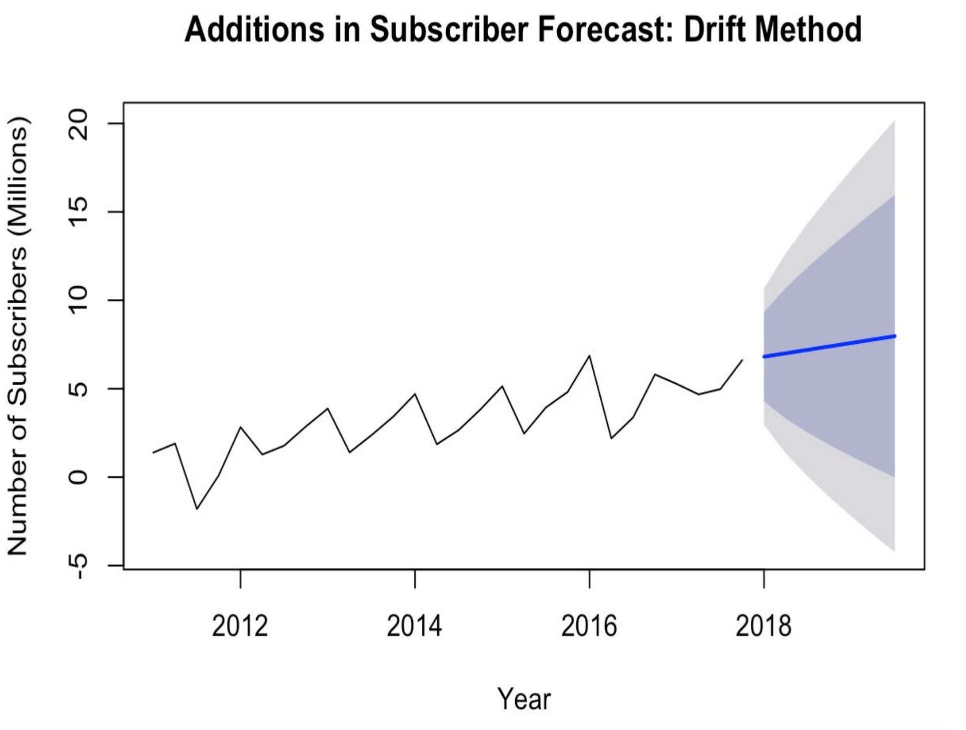
	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2018 Q1	5.27	3.508181	7.031819	2.5755306	7.964469
2018 Q2	4.68	2.918181	6.441819	1.9855306	7.374469
2018 Q3	4.98	3.218181	6.741819	2.2855306	7.674469
2018 Q4	6.62	4.858181	8.381819	3.9255306	9.314469
2019 Q1	5.27	2.778412	7.761588	1.4594449	9.080555
2019 Q2	4.68	2.188412	7.171588	0.8694449	8.490555
2019 Q3	4.98	2.488412	7.471588	1.1694449	8.790555

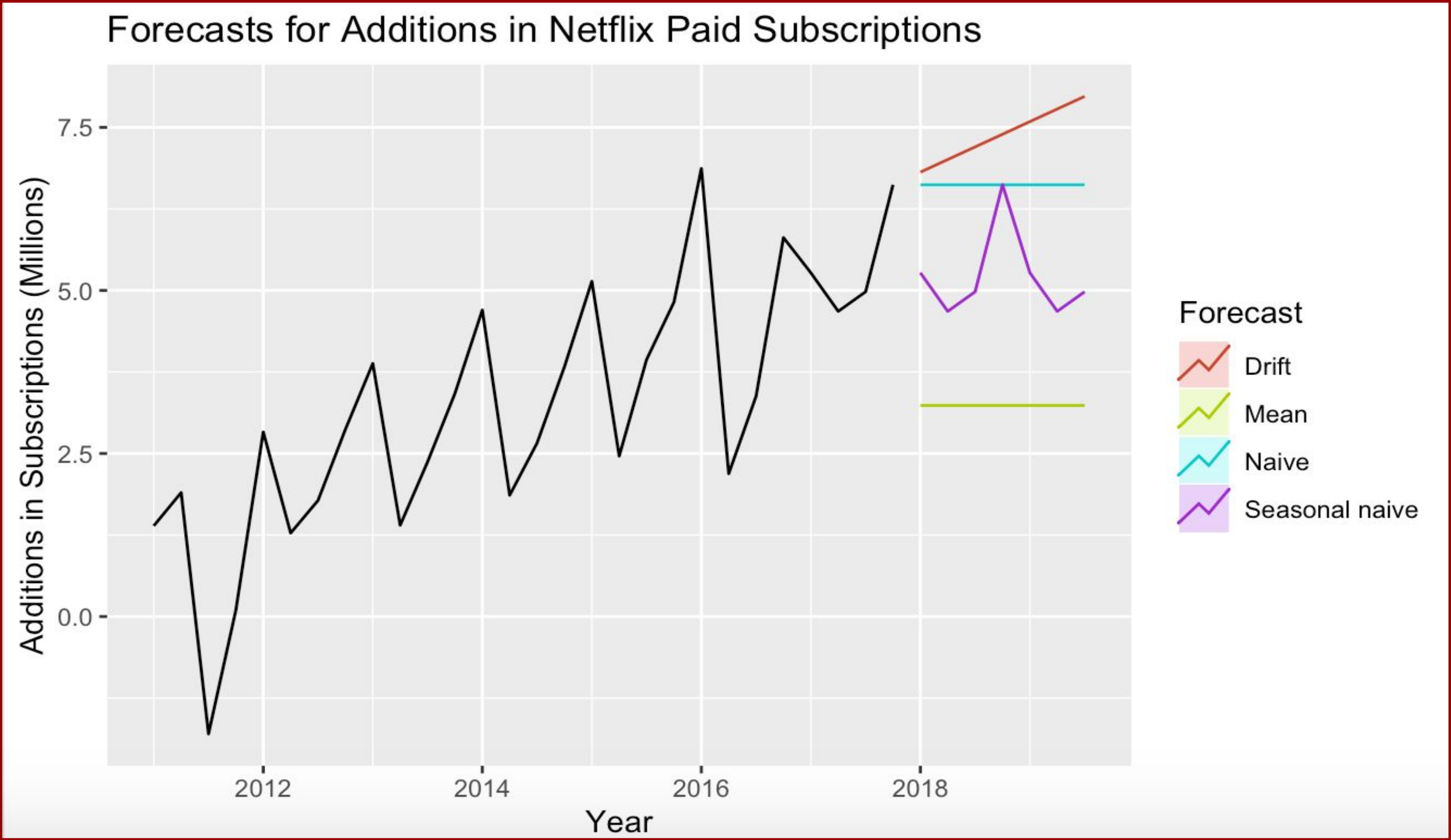


Forecasting: Drift Method

```
> #DRIFT
> driftmethod = rwf(subscribers, 7, drift = TRUE)
> driftmethod
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2018 Q1	6.813704	4.28676321	9.340644	2.94908171	10.67833
2018 Q2	7.007407	3.30834529	10.706470	1.35018008	12.66463
2018 Q3	7.201111	2.52212651	11.880096	0.04522168	14.35700
2018 Q4	7.394815	1.82570920	12.963920	-1.12239722	15.91203
2019 Q1	7.588519	1.18155499	13.995482	-2.21008664	17.38712
2019 Q2	7.782222	0.57142267	14.993022	-3.24574403	18.81019
2019 Q3	7.975926	-0.01496155	15.966813	-4.24508186	20.19693

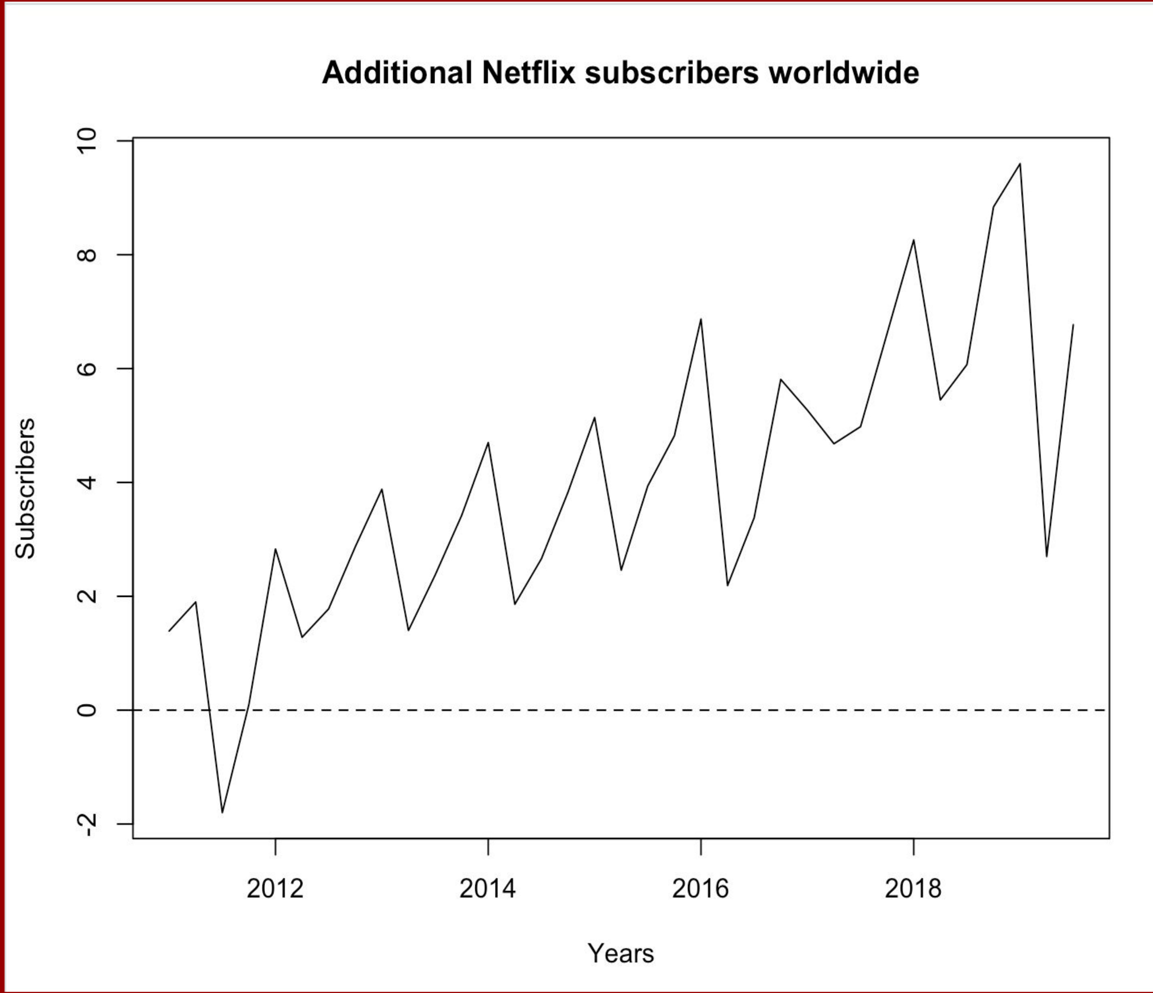



Forecasts for Additions in Netflix Paid Subscriptions
7.5
5.0
2.5
0.0
Additions in Subscriptions (Millions)
2012
2014
2016
2018
Year
Forecast
Drift
Mean
Naive
Seasonal naive

Moving Average

Purpose:

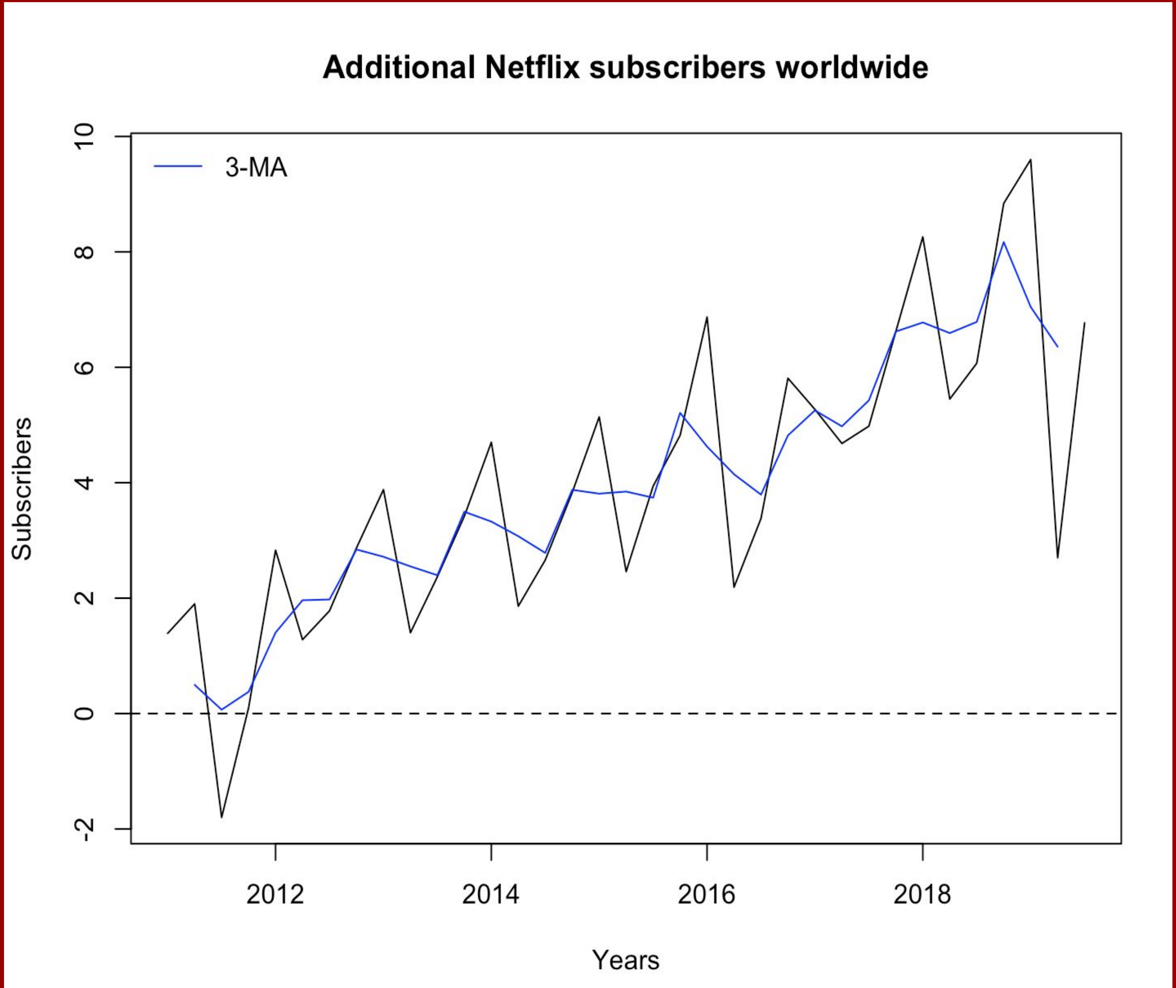
- See trend over time
- Smooths out short-term fluctuations



Moving Average

Purpose:

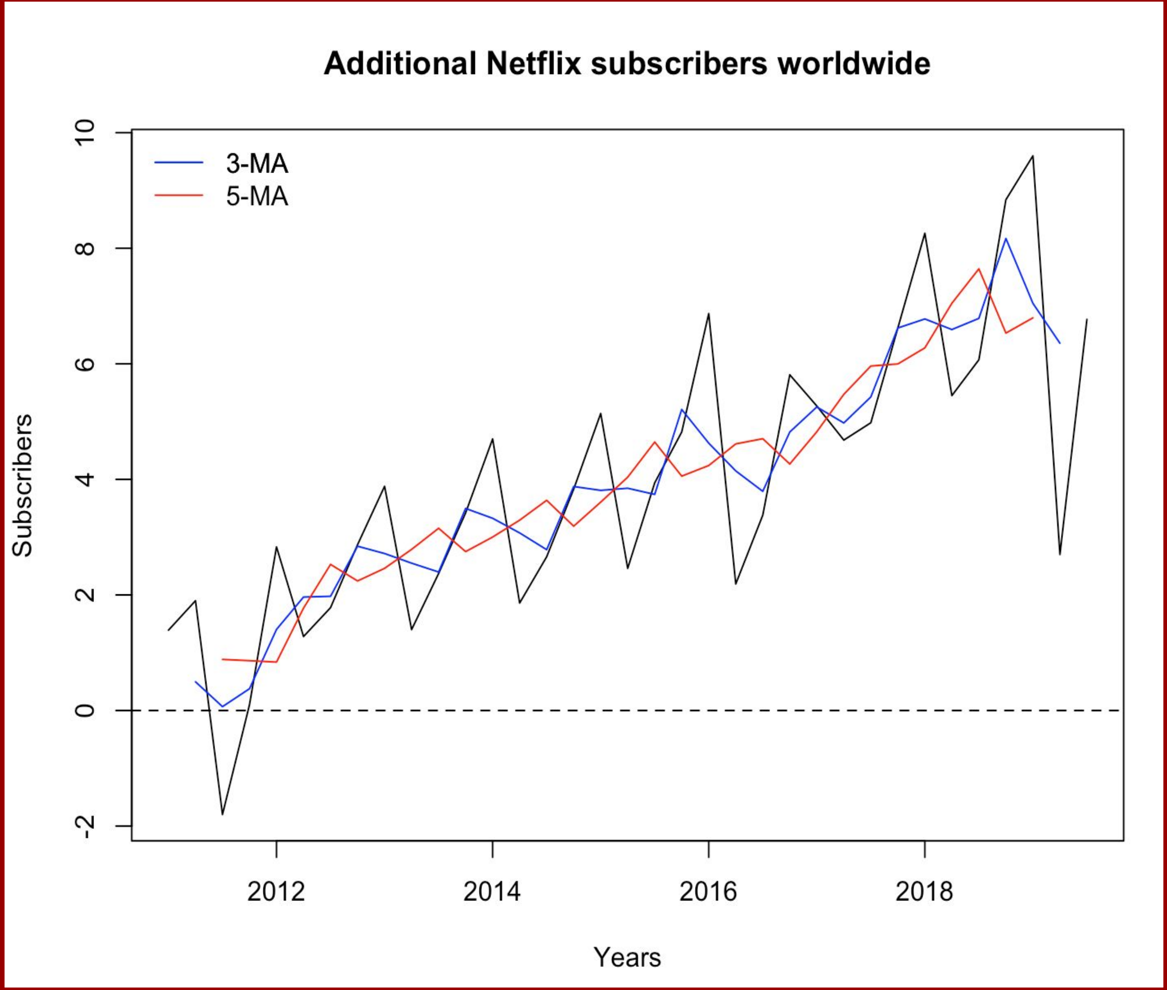
- See trend over time
- Smooths out short-term fluctuations



Moving Average

Purpose:

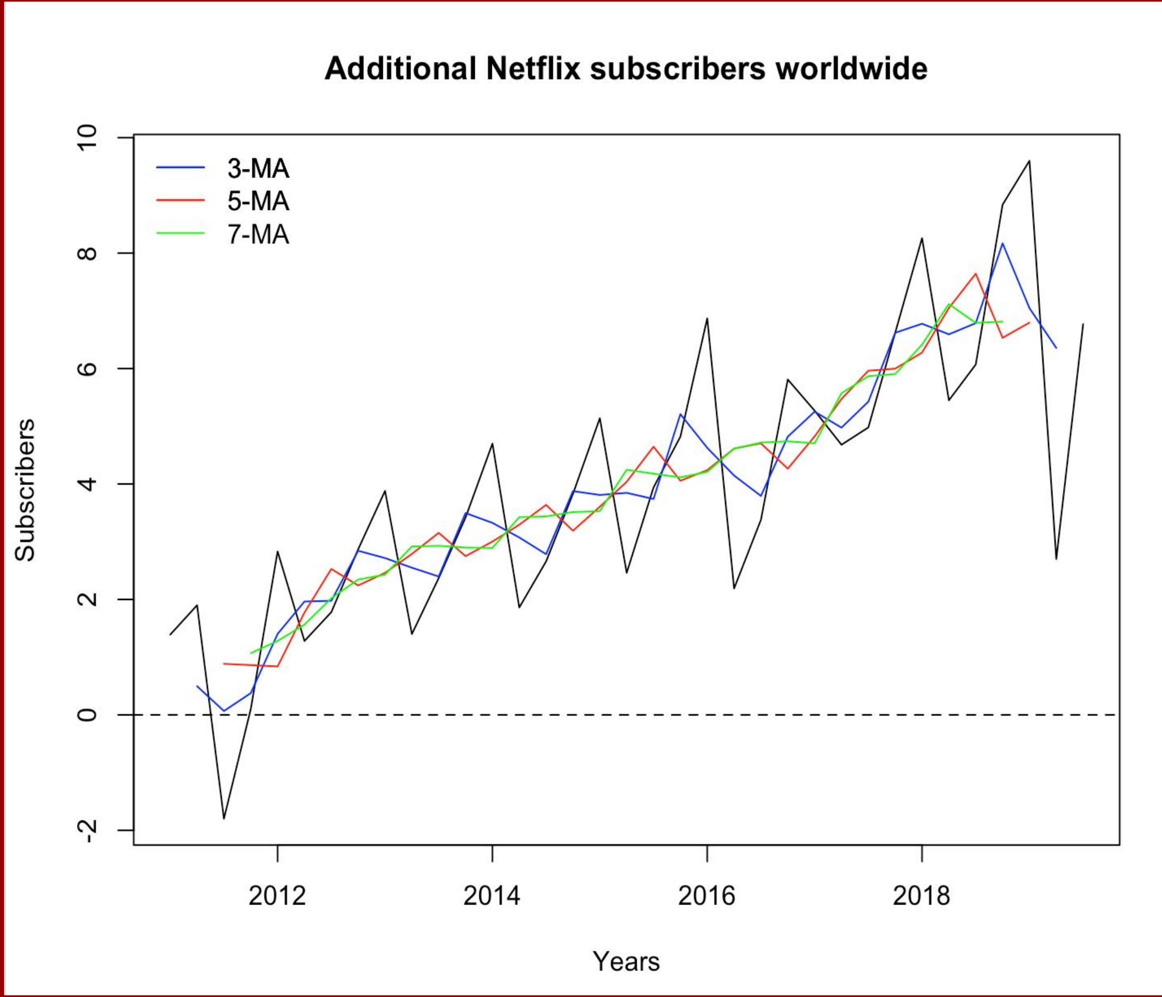
- See trend over time
- Smooths out short-term fluctuations



Moving Average

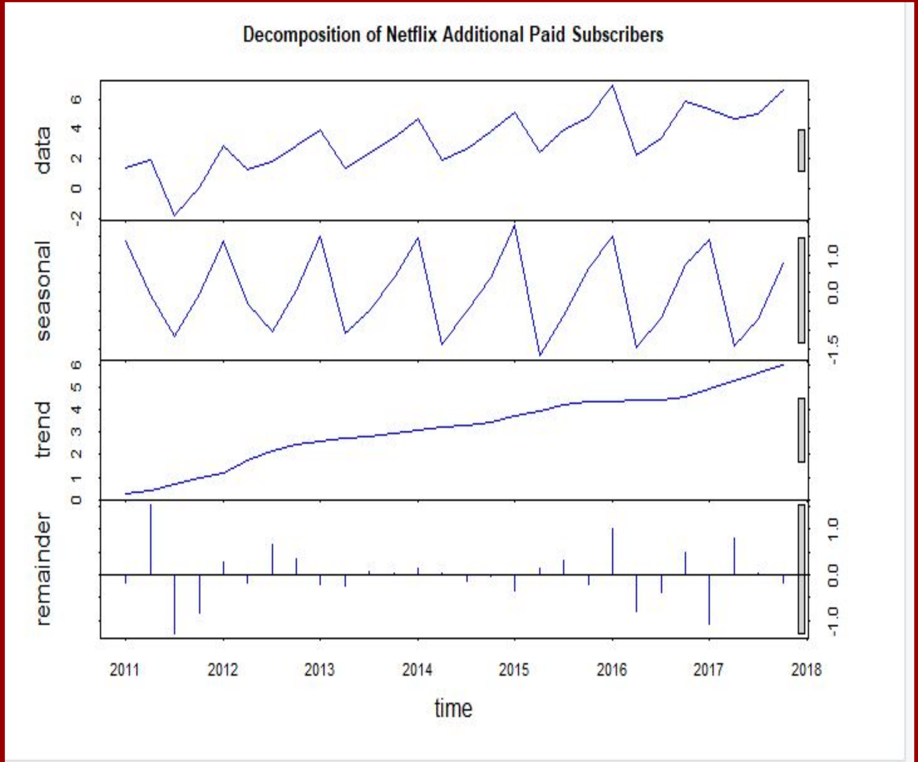
Over time:

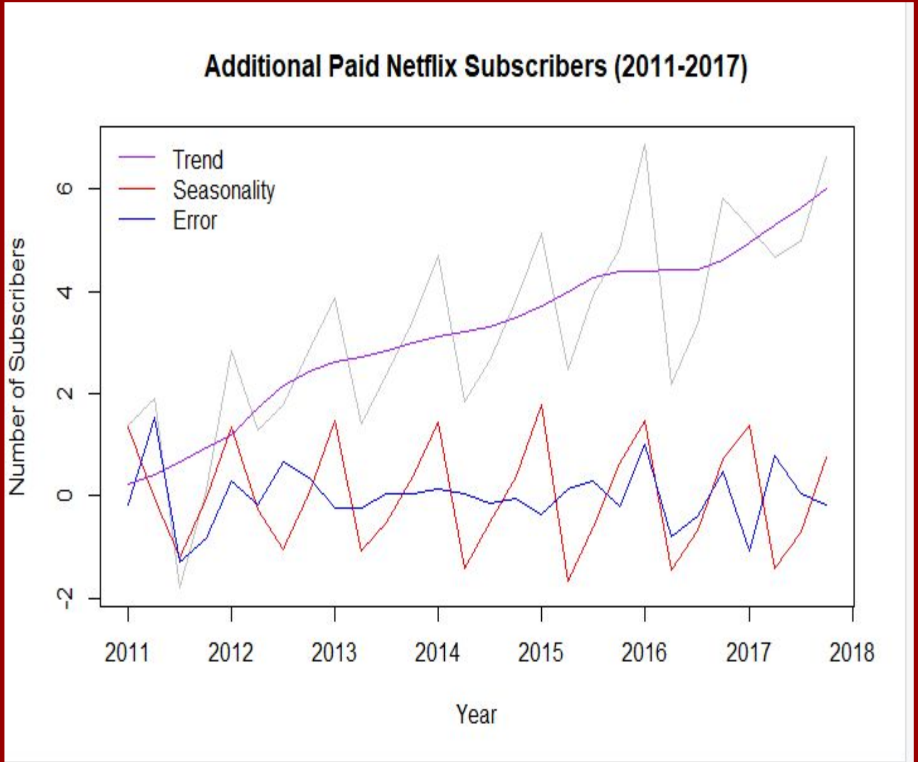
- Increasing trend
- Higher MA = smoother function



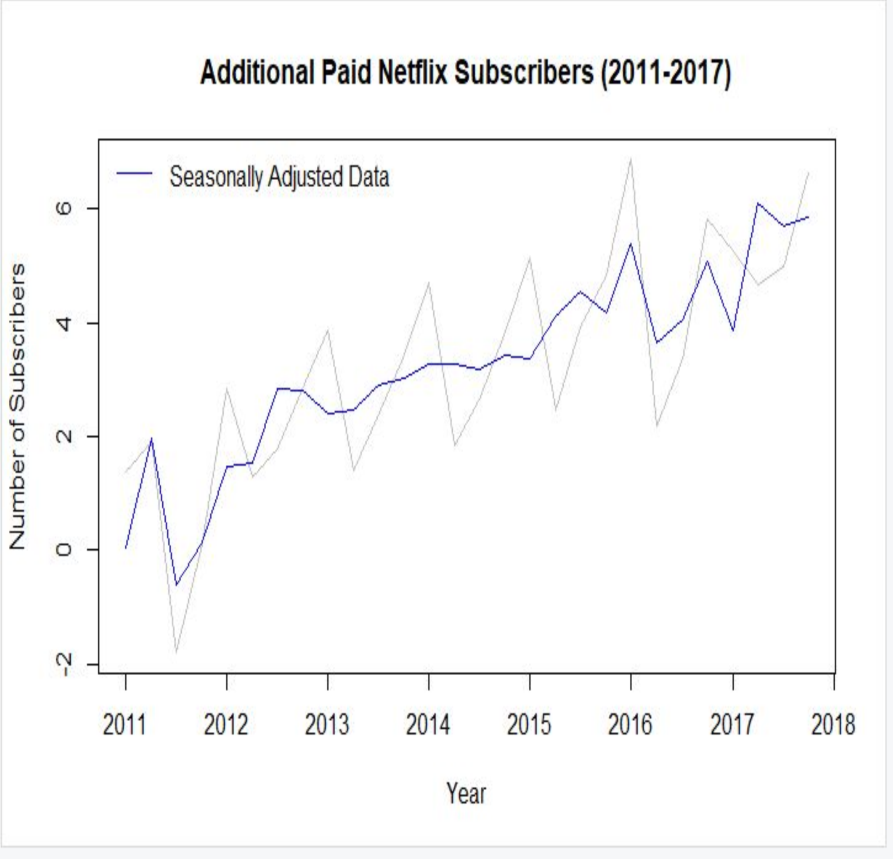
Additive Decomposition Using STL

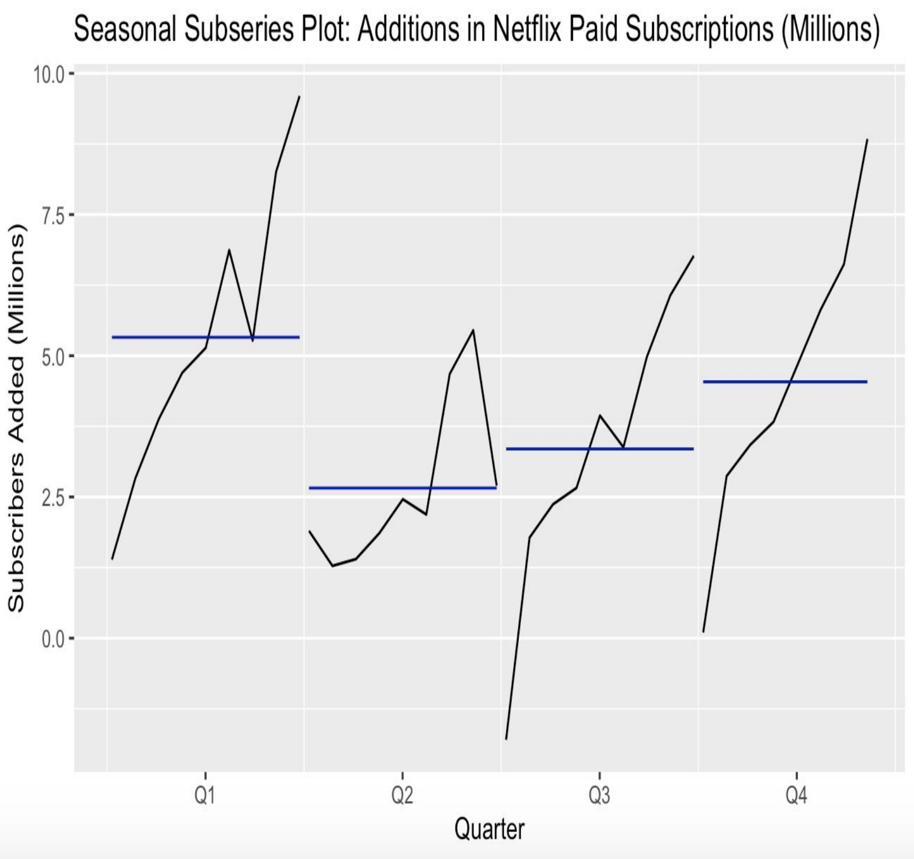
Analyzing Trend and Seasonality

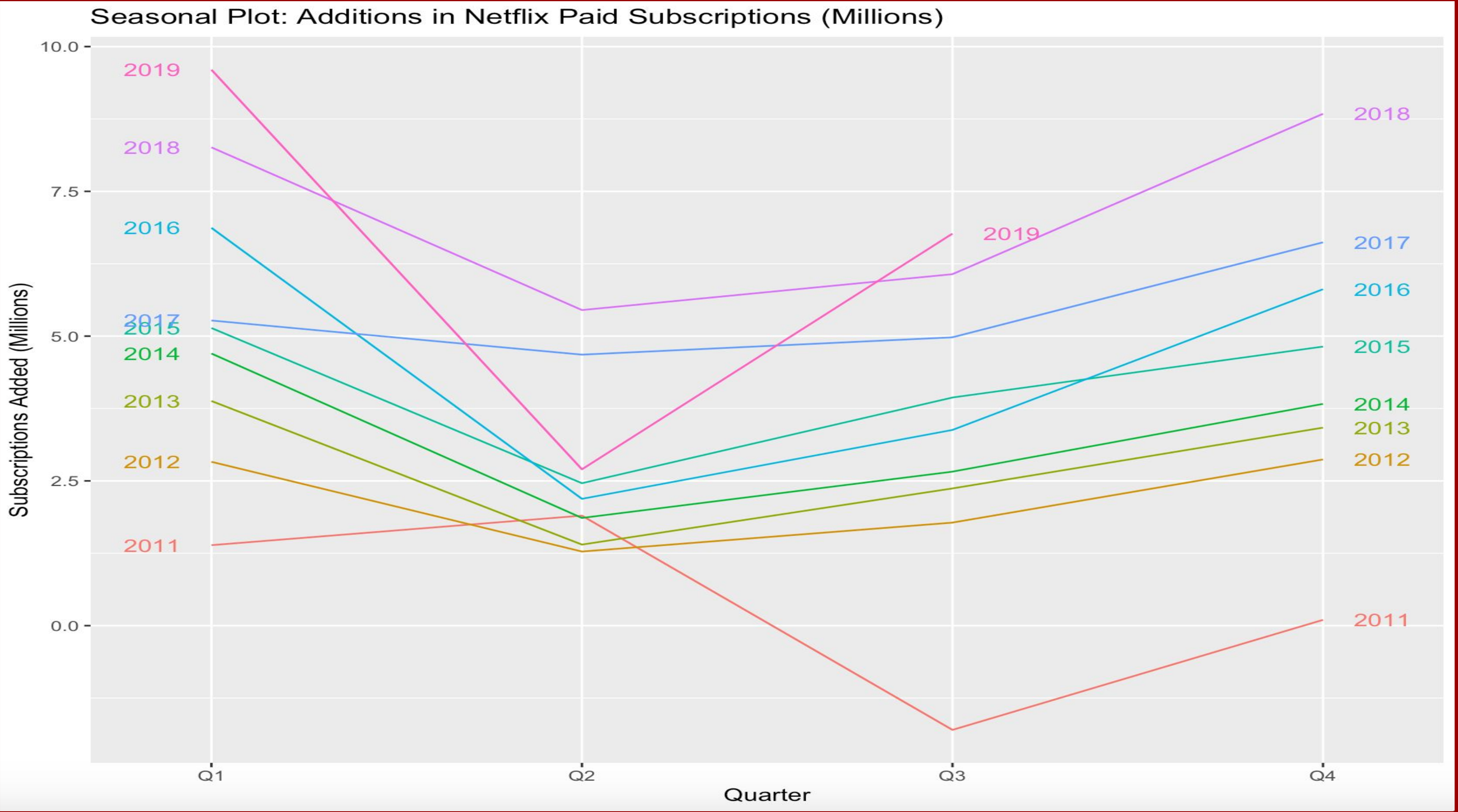




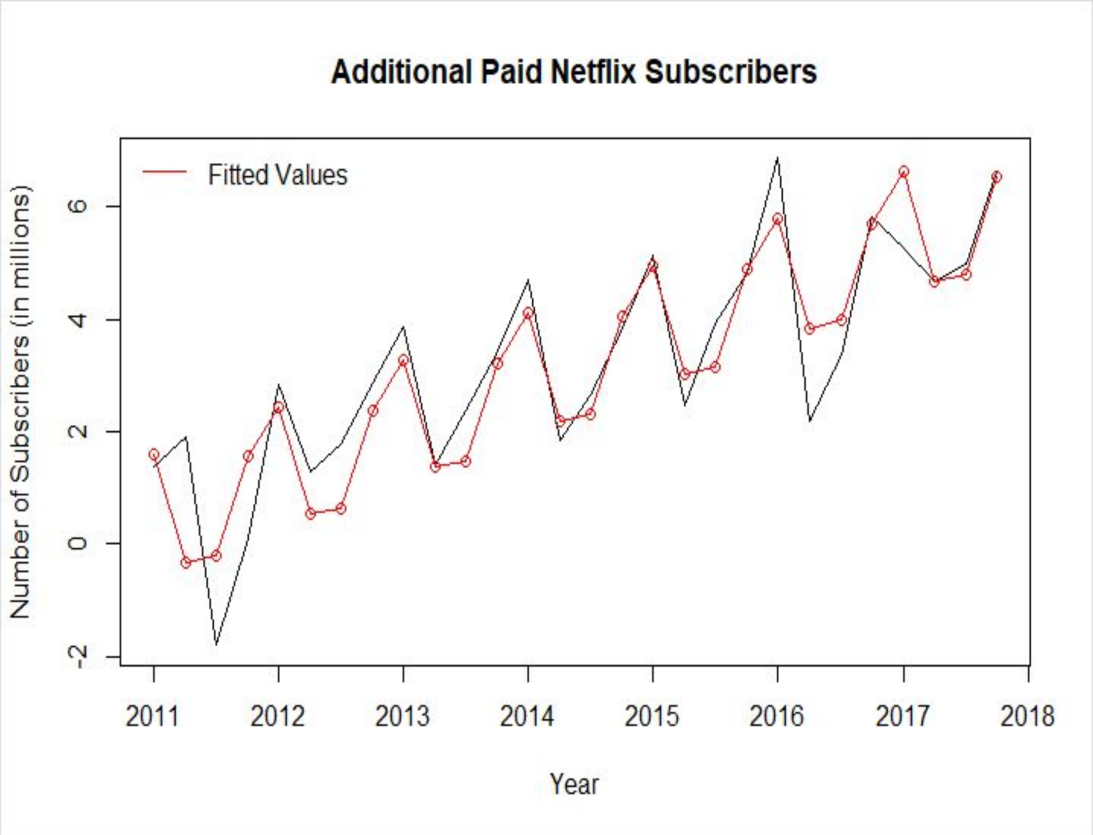
Seasonally Adjusted Data + Subseries Seasonal





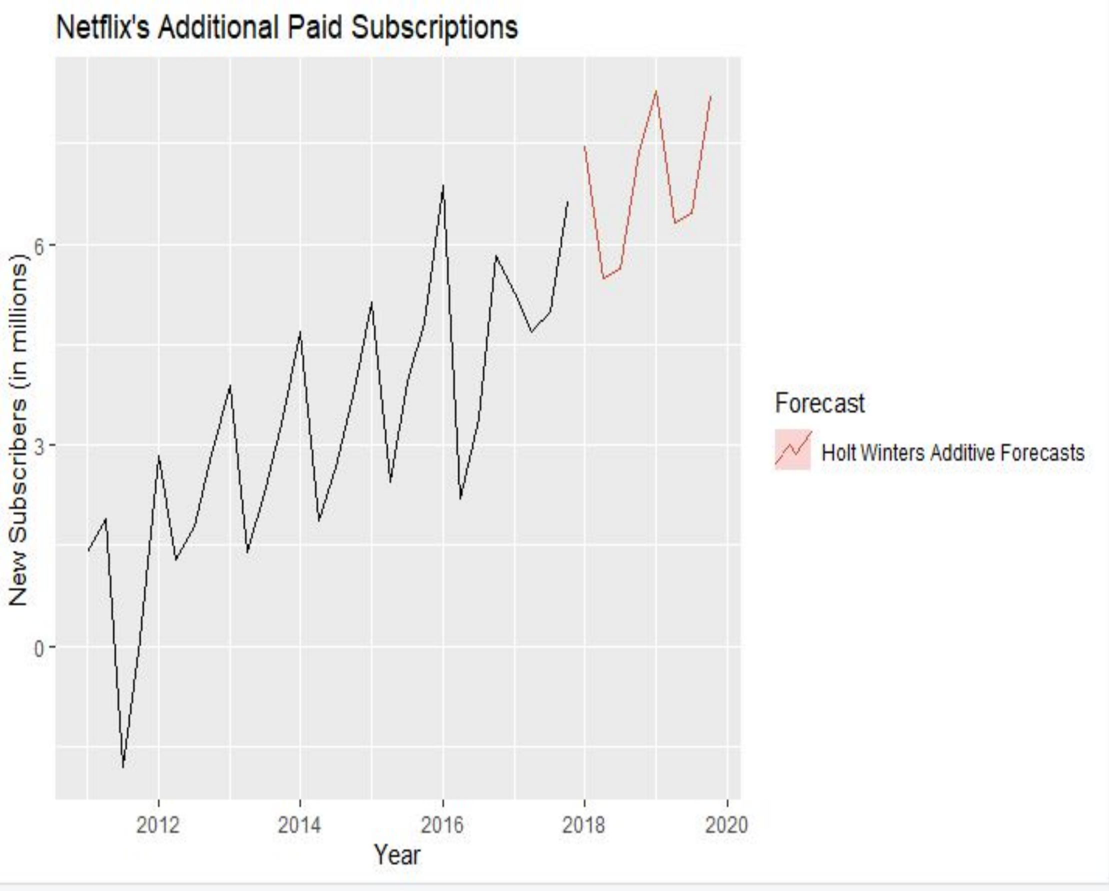

Seasonal Plot: Additions in Netflix Paid Subscriptions (Millions)
Subscriptions Added (Millions)
10.0
7.5
5.0
2.5
0.0
2019
2018
2016
2017
2015
2014
2013
2012
2011
2019
2018
2017
2016
2015
2014
2013
2012
2011
Q1
Q2
Q3
Q4
Quarter

Implementation of Holt-Winters Seasonal Method



	Qtr1	Qtr2	Qtr3	Qtr4
2011	1.6117650	-0.3262735	-0.1893177	1.5659876
2012	2.4389559	0.5280641	0.6221529	2.3794307
2013	3.2734598	1.3668904	1.4656322	3.2156925
2014	4.1111521	2.1983991	2.3068939	4.0492200
2015	4.9489359	3.0258647	3.1419753	4.8780250
2016	5.7823409	3.8510999	3.9822290	5.7085242
2017	6.6254457	4.6631389	4.8059917	6.5400591

Holt-Winters Seasonal Method Additive Forecasts



	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2018 Q1	7.440724	6.124482	8.756966	5.427705	9.453743
2018 Q2	5.493792	4.177550	6.810035	3.480773	7.506811
2018 Q3	5.638410	4.322168	6.954653	3.625391	7.651429
2018 Q4	7.371425	6.055183	8.687668	5.358406	9.384444
2019 Q1	8.271196	6.954865	9.587527	6.258042	10.284350
2019 Q2	6.324264	5.007933	7.640595	4.311110	8.337419
2019 Q3	6.468882	5.152551	7.785214	4.455727	8.482037
2019 Q4	8.201897	6.885565	9.518229	6.188742	10.215053

ARIMA

- The best ARIMA model is ARIMA(3,1,0): 3 previous observations, first differencing, and 0 previous errors

```
Coefficients:
          ar1      ar2      ar3
      -0.8094  -0.8430  -0.5815
s.e.   0.1557   0.1467   0.1538

sigma^2 estimated as 1.653:  log likelihood=-44.56
AIC=97.12   AICc=98.94   BIC=102.3
```

$$Y_t' = c - 0.8094Y_{t-1}' - 0.8430Y_{t-2}' - 0.5815Y_{t-3}'$$

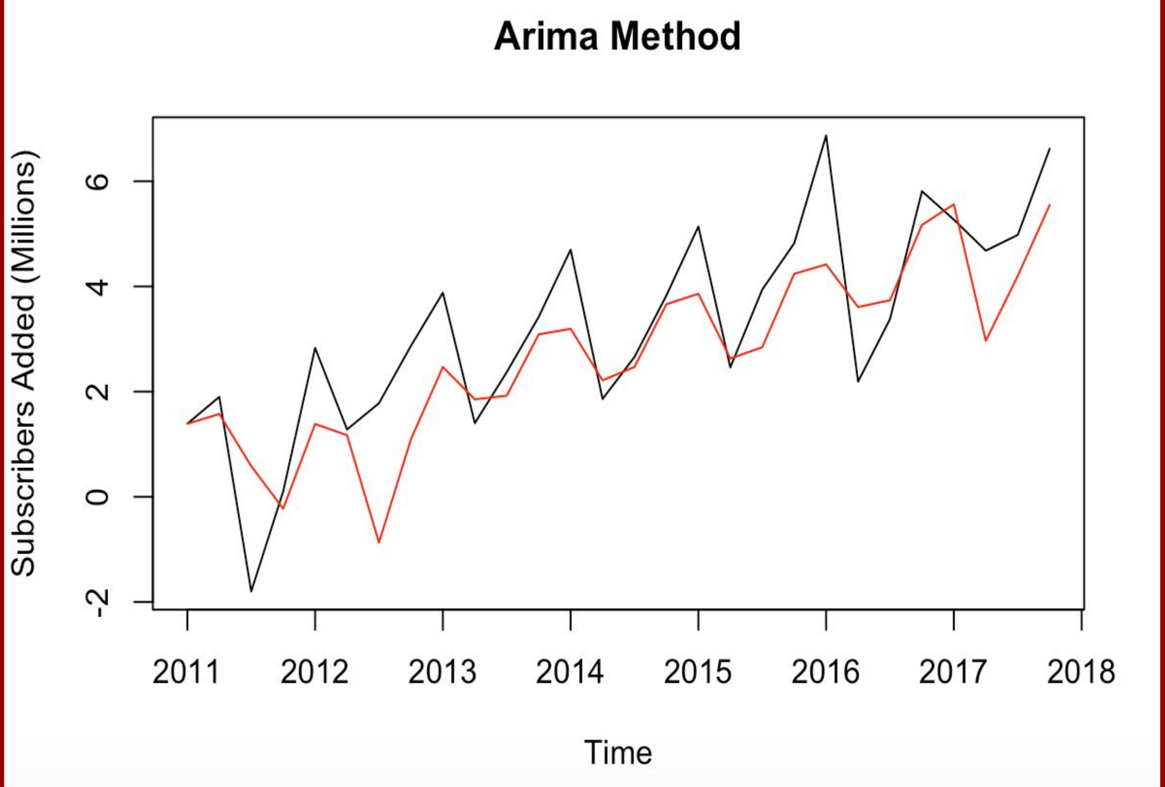
- ARIMA model with seasonality: ARIMA(0,0,0) (1,1,0)[4] with drift

```
Series: subscribers
ARIMA(0,0,0)(1,1,0)[4] with drift

Coefficients:
         sar1   drift
      -0.5225  0.1925
s.e.   0.2693  0.0364

sigma^2 estimated as 1.107:  log likelihood=-34.87
AIC=75.73   AICc=76.93   BIC=79.27
```

Implementation of ARIMA: ARIMA (3,1,0)




```
> fc=forecast(fit, h=7)
> fc
        Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2018 Q1       5.382735 3.734809 7.030661 2.862450 7.903021
2018 Q2       4.827262 3.149682 6.504842 2.261625 7.392899
2018 Q3       5.366145 3.688559 7.043731 2.800498 7.931791
2018 Q4       6.117738 4.388092 7.847384 3.472472 8.763004
2019 Q1       5.378131 3.317229 7.439033 2.226254 8.530008
2019 Q2       5.029828 2.933600 7.126055 1.823924 8.235731
2019 Q3       5.498151 3.396691 7.599610 2.284246 8.712056
```

Results

Results

Training Method	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Mean	-9.4992E-17	1.9348E+00	1.5732E+00	-1.1613E+02	1.6563E+02	1.4489E+00	4.5932E-01
Naive	1.9370E-01	1.9099E+00	1.6056E+00	7.0482E+01	1.2948E+02	1.4786E+00	-3.1836E-01
Seasonal Naive	8.3167E-01	1.3748E+00	1.0858E+00	2.5291E+01	3.4266E+01	1.0000E+00	-3.7275E-02
Drift	8.2078E-18	1.9001E+00	1.5266E+00	5.7585E+01	1.2131E+02	1.4060E+00	-3.1836E-01
Holts Winter	7.3509E-02	8.6803E-01	6.5199E-01	-4.1198E+01	7.6620E+01	4.0608E-01	-1.0463E-01
Arima	5.3143E-01	1.1905E+00	9.1903E-01	3.2028E+01	4.1966E+01	5.7240E-01	-5.1521E-02
Arima (Seasonal)	5.4783E-02	9.3259E-01	6.9481E-01	-1.0832E+00	2.7372E+01	6.3988E-01	-7.5373E-02

Testing Method	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Mean	3.5761E+00	4.1887E+00	3.7294E+00	4.4296E+01	4.9977E+01	3.4346E+00	-2.4333E-01
Naive	1.9286E-01	2.1895E+00	1.8043E+00	-1.3927E+01	3.6277E+01	1.6617E+00	-2.4333E-01
Seasonal Naive	1.6014E+00	2.4355E+00	2.1671E+00	1.3087E+01	3.4039E+01	1.9958E+00	-4.1599E-01
Drift	-5.8196E-01	2.3669E+00	1.9828E+00	-2.8349E+01	4.4009E+01	1.8261E+00	-1.6724E-01
Holts Winter	9.7329E-02	1.6039E+00	1.1453E+00	-1.1872E+01	2.6710E+01	7.1336E-01	NA
Arima	1.4414E+00	2.4332E+00	2.1071E+00	9.3036E+00	3.3958E+01	1.3124E+00	NA
Arima (Seasonal)	5.2419E-01	1.6518E+00	1.4073E+00	-3.9144E+00	2.8794E+01	1.2961E+00	-2.1838E-01

Results

Average Error	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Mean	1.8849E+00	3.0618E+00	2.6513E+00	-3.5918E+01	1.0781E+02	2.4417E+00	1.0799E-01
Naive	1.9328E-01	2.0497E+00	1.7049E+00	2.8277E+01	8.2876E+01	1.5701E+00	-2.8085E-01
Seasonal Naive	1.2165E+00	1.9051E+00	1.6265E+00	1.9189E+01	3.4153E+01	1.4979E+00	-2.2663E-01
Drift	-2.9098E-01	2.1335E+00	1.7547E+00	1.4618E+01	8.2661E+01	1.6160E+00	-2.4280E-01
Holts Winter	8.5419E-02	1.2360E+00	8.9867E-01	-2.6535E+01	5.1665E+01	5.5972E-01	NA
Arima	9.8643E-01	1.8118E+00	1.5131E+00	2.0666E+01	3.7962E+01	9.4239E-01	NA
Arima (Seasonal)	2.8949E-01	1.2922E+00	1.0511E+00	-2.4988E+00	2.8083E+01	9.6798E-01	-1.4688E-01

- **Holt Winters Seasonal Method** has the greatest number of minimum errors
- The Mean, Naive, Seasonal Naive, Drift, and Arima models have the greatest errors

Conclusion

Conclusions

- Overall, the best forecasting method is the **Holt's Winter Seasonal Method.**
- Takes seasonality and trend into account.
    - Netflix has certain seasons of growth, but also an overall increasing trend.
- Can be used to predict future number of paid subscribers added each quarter.
- This is evidenced by the fact that Holt's Winter Seasonal Method has the lowest values (closest to 0), for:
    - ME
    - RMSE
    - MAE
    - MASE
- This indicates that the model has the lowest errors, which are closest to 0, indicating the closest fit.

References

https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/ (Data Source)

https://www.statista.com/chart/16684/netflix-subscription-prices-in-the-united-states/

https://otexts.com/fpp2/moving-averages.html

https://www.statista.com/chart/16684/netflix-subscription-prices-in-the-united-states/

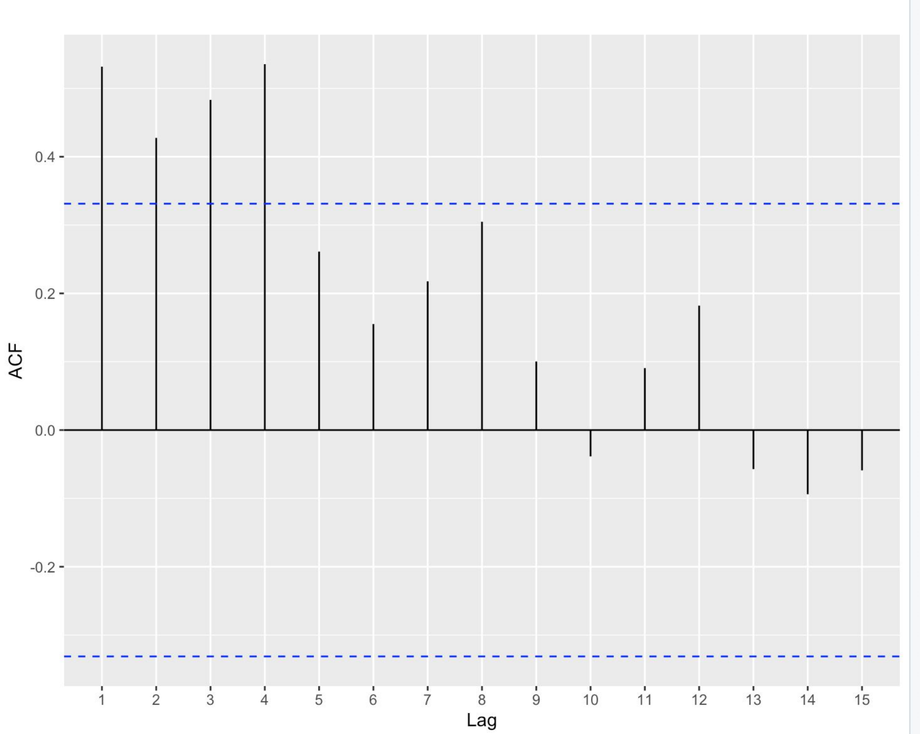
https://www.vox.com/2019/1/16/18185174/netflix-price-increase-subscription-chart-original-content-streaming

https://rcompanion.org/handbook/G_14.html

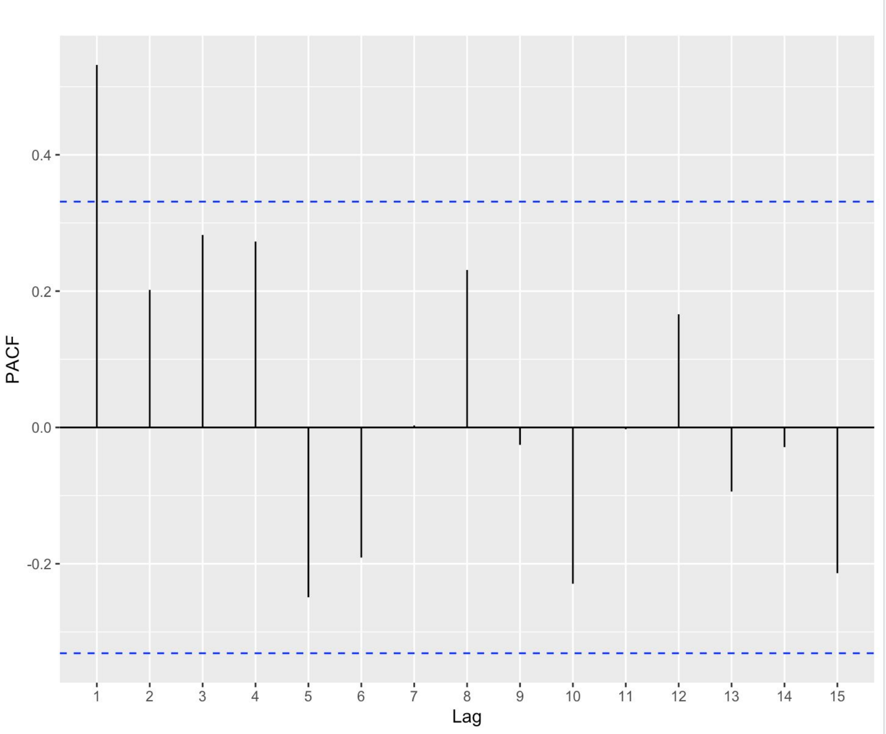
https://www.forbes.com/sites/theyec/2011/12/28/5-business-lessons-from-the-netflix-pricing-debacle/#4c18ff64d2a7

Appendix

ACF vs. PACF



ACF
0.4
0.2
0.0
-0.2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Lag



PACF
0.4
0.2
0.0
-0.2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Lag

Challenges

- Somewhat small dataset
    - Q1 2011 - Q3 2019 (35 different data points total)
    - Difficult to attain data before 2011 because of changes in reporting