

Time Series Student Project  
 Name: Reid Earnhardt  
 Course: Fall 2015 TS

**Title: Percent of Men with full beards, 1866 – 1911**

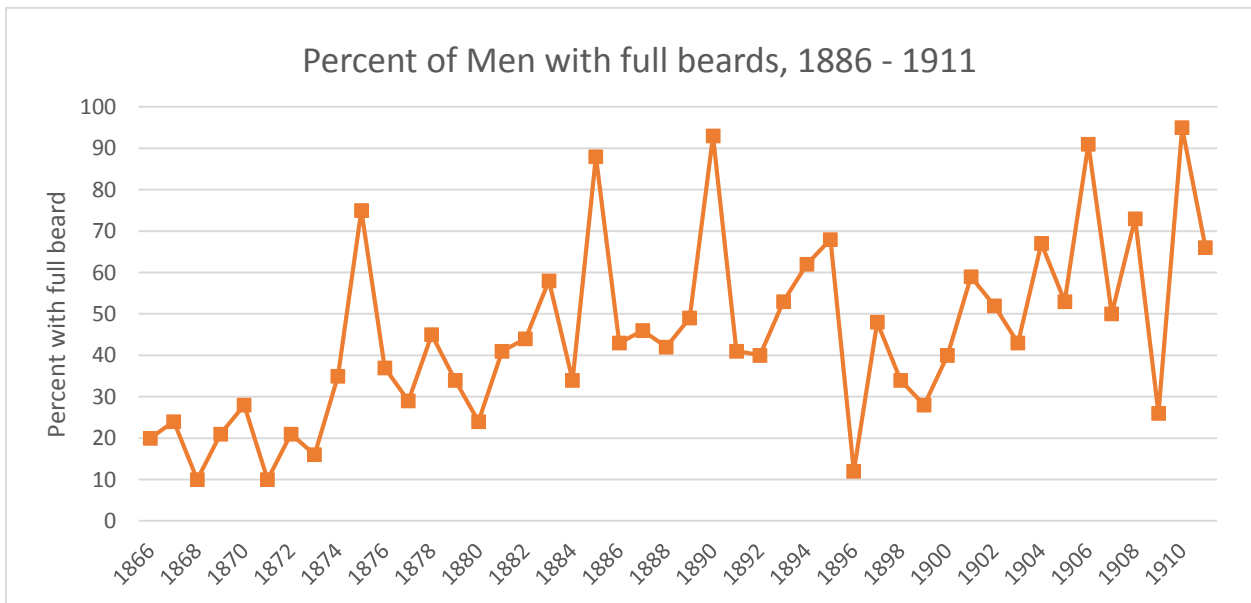
Introduction

Some may say a man wearing a full beard is back in style, and has been increasing in popularity among hipster identifying cultures. This project looks at beard style trends not from today, but over a century ago, and aims to fit an ARIMA model to the time series using methods learned in the time series on-line course.

Data Source

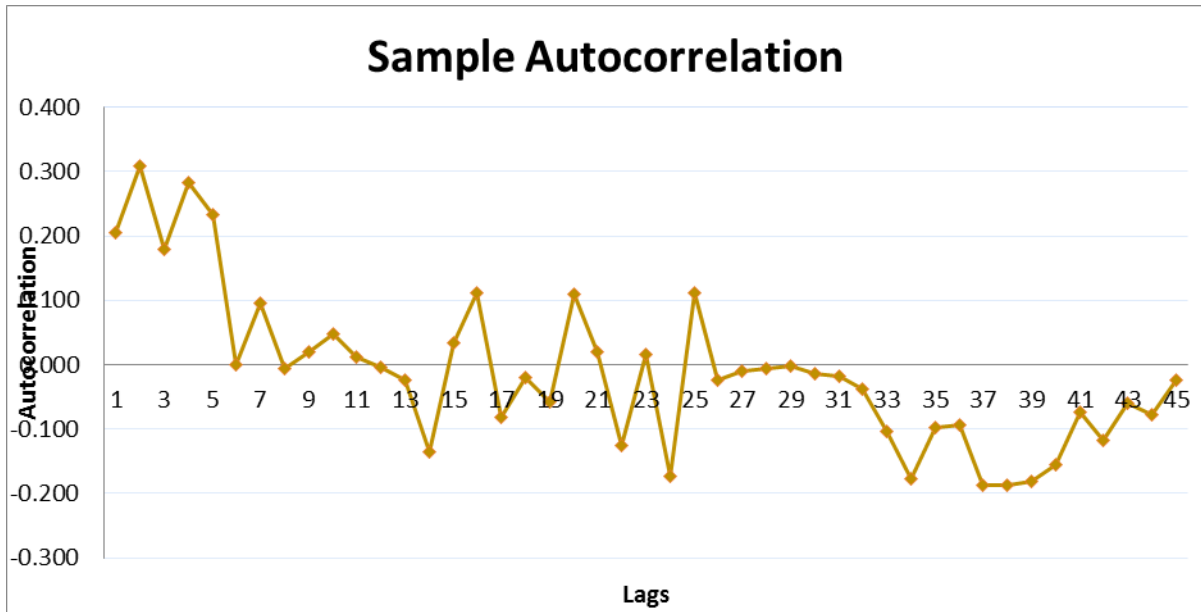
Data is sourced from datamarket.com and is presented below.

<u>Year</u>	<u>Percent with Full Beard</u>	<u>Year</u>	<u>Percent with Full Beard</u>	<u>Year</u>	<u>Percent with Full Beard</u>	<u>Year</u>	<u>Percent with Full Beard</u>
1866	20	1878	45	1890	93	1902	52
1867	24	1879	34	1891	41	1903	43
1868	10	1880	24	1892	40	1904	67
1869	21	1881	41	1893	53	1905	53
1870	28	1882	44	1894	62	1906	91
1871	10	1883	58	1895	68	1907	50
1872	21	1884	34	1896	12	1908	73
1873	16	1885	88	1897	48	1909	26
1874	35	1886	43	1898	34	1910	95
1875	75	1887	46	1899	28	1911	66
1876	37	1888	42	1900	40		
1877	29	1889	49	1901	59		



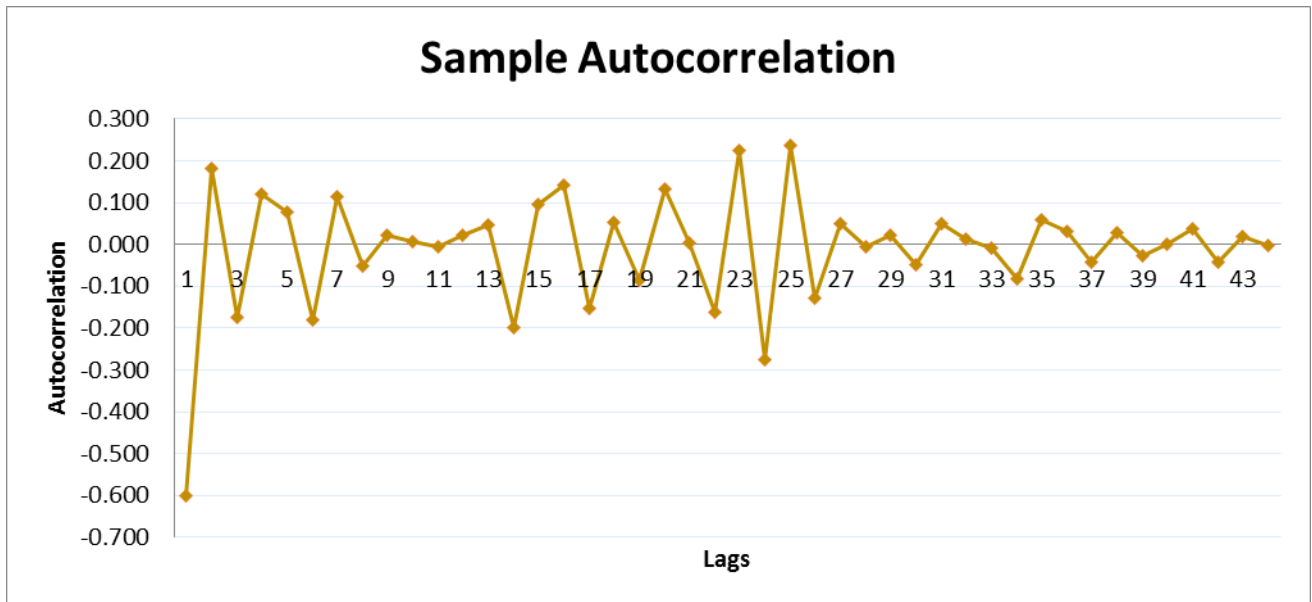
Stationary

The plot of the time series would seem to indicate an upward trend over the period. I hypothesized that the data is not stationary. To test this hypothesis, I graphed the sample autocorrelation over all lags.

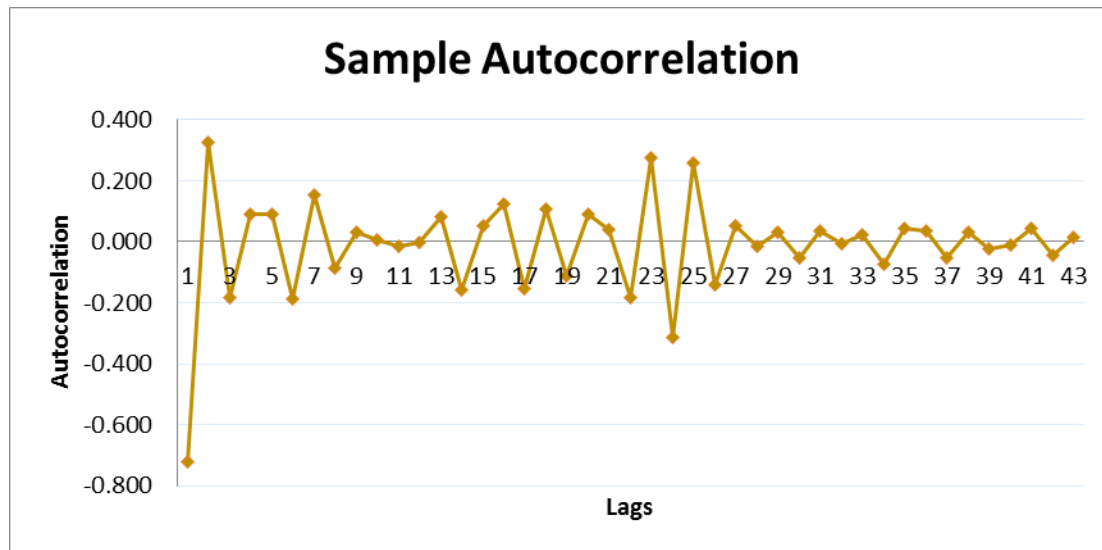


This confirms the time series is not stationary since the values do not drop quickly to zero, and they do not return to zero at subsequent lags with minimum fluctuations.

The correlogram of the first difference of  $Y(t)$  can be seen below.



The correlogram of the second difference of  $Y(t)$  can be seen below.



Both first and second differences drop quickly to zero and oscillate around zero. The oscillations in both graphs fluctuate minimally towards the end of the lags. Both suggest stationarity and since second differences do not present any significant improvement over first differences, I will choose first differences to make the process stationary in order to avoid over differencing.

#### Autoregressive Process

We will now use  $ARI(1,1)$ ,  $ARI(2,1)$  and  $ARI(3,1)$  processes to fit a model to the time series.

The  $ARI(1,1)$  model is:

$$Y_t - Y_{t-1} = \phi_1(Y_{t-1} - Y_{t-2}) + \theta$$

The  $ARI(2,1)$  model is:

$$Y_t - Y_{t-1} = \phi_1(Y_{t-1} - Y_{t-2}) + \phi_2(Y_{t-2} - Y_{t-3}) + \theta$$

The  $ARI(3,1)$  model is:

$$Y_t - Y_{t-1} = \phi_1(Y_{t-1} - Y_{t-2}) + \phi_2(Y_{t-2} - Y_{t-3}) + \phi_3(Y_{t-3} - Y_{t-4}) + \theta$$

Using regression analysis in Excel, I obtained the following results:

$$ARI(1,1): Y_t = -0.6181(Y_{t-1} - Y_{t-2}) + Y_{t-1} + 2.0524$$

$$ARI(2,1): Y_t = -0.7738(Y_{t-1} - Y_{t-2}) - 0.2810(Y_{t-2} - Y_{t-3}) + Y_{t-1} + 2.3494$$

$$ARI(3,1): Y_t = -0.8634(Y_{t-1} - Y_{t-2}) - 0.5416(Y_{t-2} - Y_{t-3}) - 0.3825(Y_{t-3} - Y_{t-4}) + Y_{t-1} + 2.9838$$

Please see appendix A for details of the regression analysis.

## Regression Summary

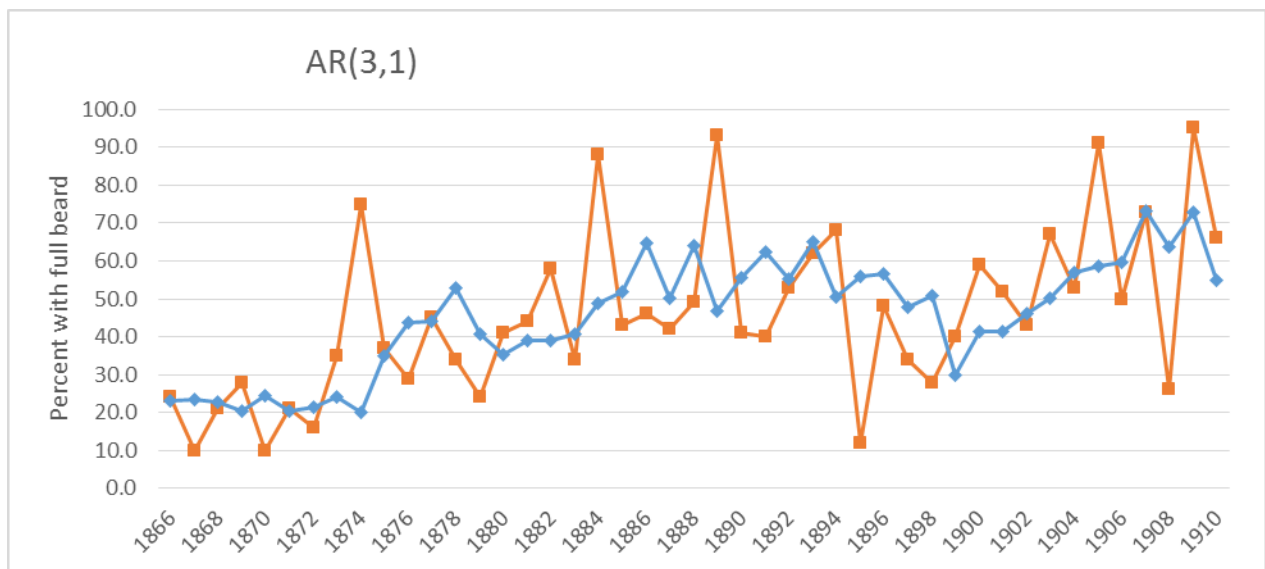
	R <sup>2</sup>	Adj. R <sup>2</sup>	Std. Error
ARI(1,1)	0.37149	0.35687	22.06164
ARI(2,1)	0.41370	0.38578	21.56016
ARI(3,1)	0.48636	0.44878	20.42453

## Residual Diagnostic

I will use the Box Pierce Q Statistic to determine if the residuals from the ARIMA model are a random walk. See appendix A for details of the analysis. Using a critical Chi2 value for a 10% significance level, the BPQS calculated is below the critical value after a few lags. We do not reject the null hypothesis that the residuals are a white noise process.

## Conclusion

The time series of historical percent of men with full beards is not a stationary series. In order to make it stationary it was transformed by taking first differences. Autoregressive models were then considered of orders 1, 2 and 3. Regression analysis produced three ARIMA models and further residual analysis confirmed residuals of each model may follow a white noise process. Based on these results I choose the ARI(3,1) model to fit the time series due to it having the largest R<sup>2</sup> values and lowest standard error.



As shown, the fitted line in blue follows the time series on its general course but has some obvious deviations from actual. Due to the dates in the series and unavailable information on how the data was collected, some of the collection values may be subject. I notice at least 5 or 6 outlying observations and wonder how accurate those data points are. Nonetheless, if you are or know somebody thinking about growing a beard, know you are in good company as many men over the centuries have embraced the fashion as their own.

## Appendix A

Time series, differences and fitted models:

Percent of Men with full beards, 1866 – 1911					
Year	Percent with Full Beard	Percent with Full Beard $Y(t) - Y(t-1)$	Percent with Full Beard $Y(t-1) - Y(t-2)$	Percent with Full Beard $Y(t-2) - Y(t-3)$	Percent with Full Beard $Y(t-3) - Y(t-4)$
1866	24.0	4.0	0.0	0.0	0.0
1867	10.0	-14.0	4.0	0.0	0.0
1868	21.0	11.0	-14.0	4.0	0.0
1869	28.0	7.0	11.0	-14.0	4.0
1870	10.0	-18.0	7.0	11.0	-14.0
1871	21.0	11.0	-18.0	7.0	11.0
1872	16.0	-5.0	11.0	-18.0	7.0
1873	35.0	19.0	-5.0	11.0	-18.0
1874	75.0	40.0	19.0	-5.0	11.0
1875	37.0	-38.0	40.0	19.0	-5.0
1876	29.0	-8.0	-38.0	40.0	19.0
1877	45.0	16.0	-8.0	-38.0	40.0
1878	34.0	-11.0	16.0	-8.0	-38.0
1879	24.0	-10.0	-11.0	16.0	-8.0
1880	41.0	17.0	-10.0	-11.0	16.0
1881	44.0	3.0	17.0	-10.0	-11.0
1882	58.0	14.0	3.0	17.0	-10.0
1883	34.0	-24.0	14.0	3.0	17.0
1884	88.0	54.0	-24.0	14.0	3.0
1885	43.0	-45.0	54.0	-24.0	14.0
1886	46.0	3.0	-45.0	54.0	-24.0
1887	42.0	-4.0	3.0	-45.0	54.0
1888	49.0	7.0	-4.0	3.0	-45.0
1889	93.0	44.0	7.0	-4.0	3.0
1890	41.0	-52.0	44.0	7.0	-4.0
1891	40.0	-1.0	-52.0	44.0	7.0
1892	53.0	13.0	-1.0	-52.0	44.0
1893	62.0	9.0	13.0	-1.0	-52.0
1894	68.0	6.0	9.0	13.0	-1.0
1895	12.0	-56.0	6.0	9.0	13.0
1896	48.0	36.0	-56.0	6.0	9.0
1897	34.0	-14.0	36.0	-56.0	6.0
1898	28.0	-6.0	-14.0	36.0	-56.0
1899	40.0	12.0	-6.0	-14.0	36.0
1900	59.0	19.0	12.0	-6.0	-14.0
1901	52.0	-7.0	19.0	12.0	-6.0
1902	43.0	-9.0	-7.0	19.0	12.0
1903	67.0	24.0	-9.0	-7.0	19.0
1904	53.0	-14.0	24.0	-9.0	-7.0
1905	91.0	38.0	-14.0	24.0	-9.0
1906	50.0	-41.0	38.0	-14.0	24.0
1907	73.0	23.0	-41.0	38.0	-14.0
1908	26.0	-47.0	23.0	-41.0	38.0
1909	95.0	69.0	-47.0	23.0	-41.0
1910	66.0	-29.0	69.0	-47.0	23.0





