

# Exploring Climate Temperature Data - Part 1

---

Exploring climate temperature data with Kozo and IRIS



## Introduction

---

This main goal for this document is to demonstrate Kozo as a tool for exploring and writing dynamic data rich documents.

## The Data

---

It looks like the following three datasets have a lot of raw data to work with.

- NOAA MLOST
- NASA GISTEMP
- Met Office HadCrut

But I soon find out that the data is not in a friendly format.

I eventually found a curated dataset from Berkeley Earth which is affiliated with Lawrence Berkeley National Laboratory.

<https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data>

I selected the largest dataset I could find. A 520mb CSV files containing surface temperatures by City. On inspection it contains 8.6 million records. This should be good enough for now.

I import the data into an IRIS table, making sure I have adequate indexes. I also decide to break the dates down into separate Year and Month fields as this will make slicing the data up by month much faster.

# Introduction

---

This main goal for this document is to demonstrate Kozo as a tool for exploring and writing dynamic data rich documents.

## The Data

---

It looks like the following three datasets have a lot of raw data to work with.

- NOAA MLOST
- NASA GISTEMP
- Met Office HadCrut

But I soon find out that the data is not in a friendly format.

I eventually found a curated dataset from Berkeley Earth which is affiliated with Lawrence Berkeley National Laboratory.

<https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data>

I selected the largest dataset I could find. A 520mb CSV files containing surface temperatures by City. On inspection it contains 8.6 million records. This should be good enough for now.

I import the data into an IRIS table, making sure I have adequate indexes. I also decide to break the dates down into separate Year and Month fields as this will make slicing the data up by month much faster.

# Years

---

It's at this point I wonder how complete the range of dates in the data are. Since I can do this within my Kozo report I add in a simple SQL query and then output the results using Moustache template syntax.

1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760  
1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778  
1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796  
1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814  
1815 1816 1817 1818 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832  
1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850  
1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868  
1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886  
1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904  
1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922  
1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940  
1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958  
1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976  
1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994  
1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012  
2013

Now lets see how many temperature readings I have per year. I only need to look at the first year of each decade so I narrow the SQL query down to unclude just these.

## Temperature Counts By Year

---

Year	Total
2010	42120
2000	42120
1990	42120
1980	42120
1970	42120
1960	42120
1950	42120
1940	42120

Year	Total
1930	42120
1920	42120
1910	42120
1900	42120
1890	42120
1880	41892
1870	41796
1860	41532
1850	37848
1840	30492
1830	25308
1820	19362
1810	16056
1800	15660
1790	10428
1780	10251
1770	9480
1760	9360
1750	8484

It looks as if there is a consistant 42,000 temperature readings per year for the past 170 years. The number of readings then starts to tail off from the mid 1800's to the first records that started in the mid 1700's.

At this point I have a couple of observations. The Berkley data file I am using is already three years out of date. It was currated I think in 2019 and the data it includes only goes up until 2013, which is a good ten years behind current readings.

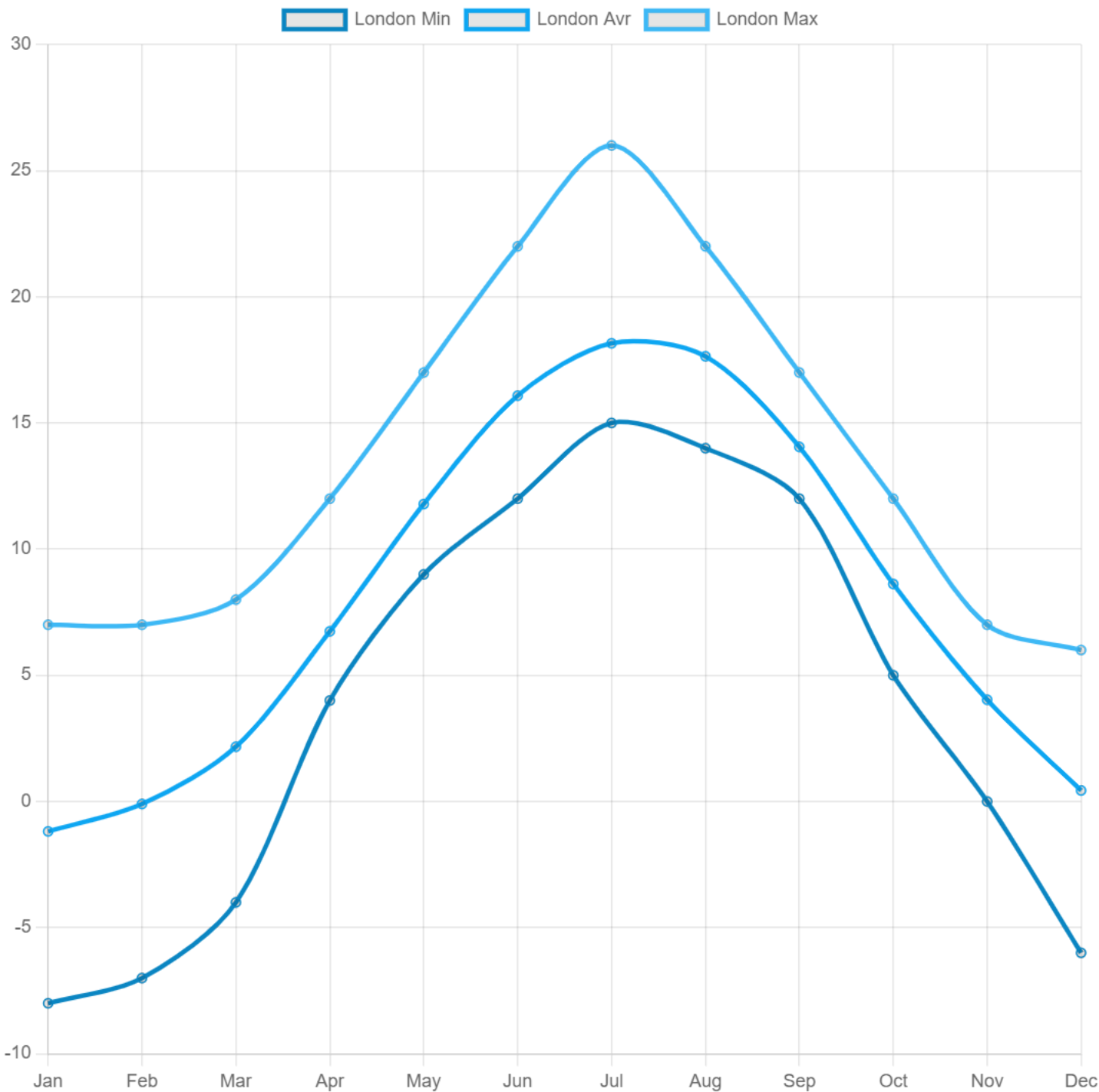
## Time for some graphs

To begin with I think I want to just explore plain old temperature line graphs over time. Although I realise I potentially have 270 years x 12 months of data per city, thats 3240 data points which is going

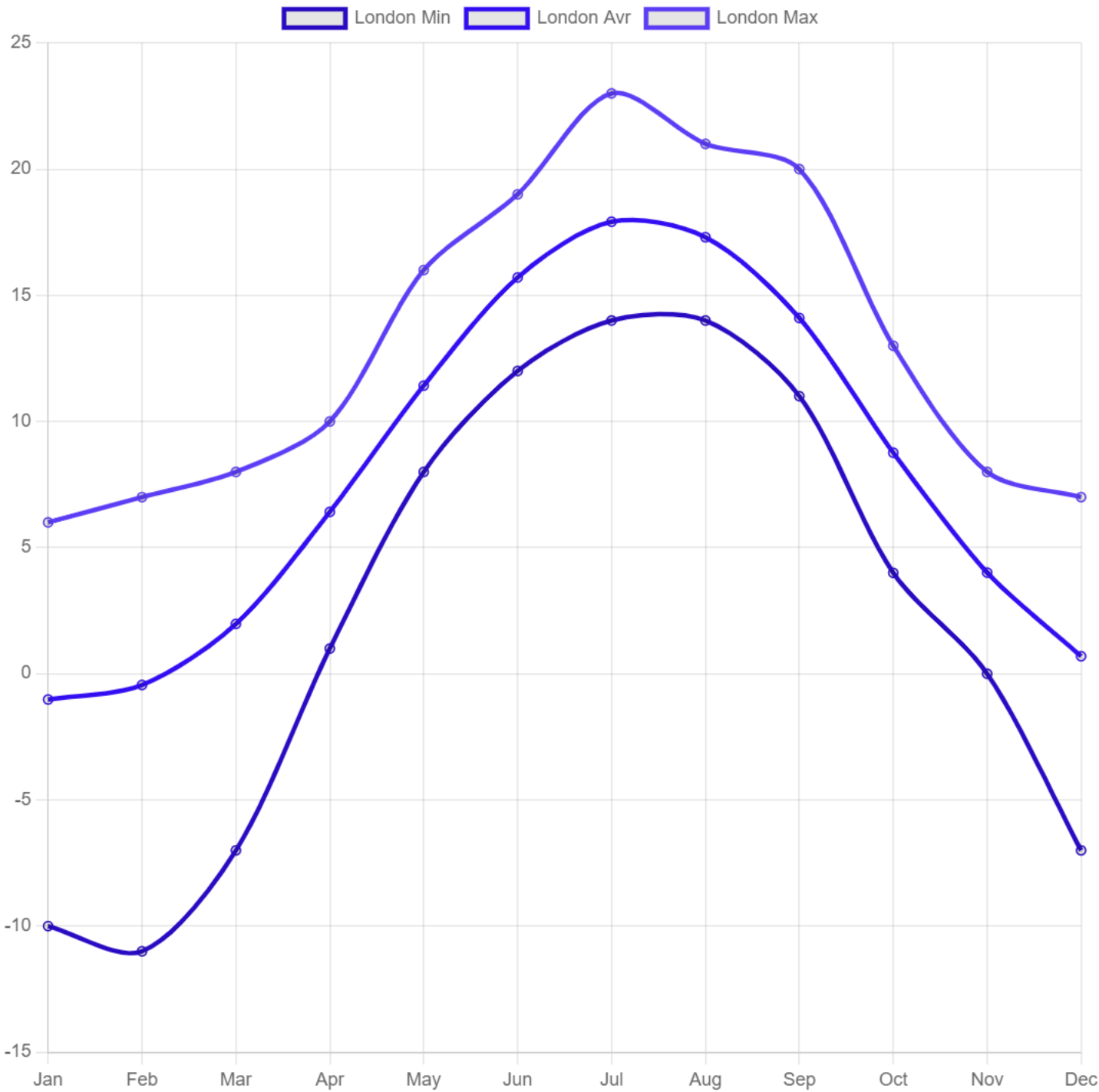
to look really compressed and spikey considering temperatures by month will jump around. This needs some thinking about.

After considering this for a while, I want to start by getting a feel for the upper and lower temperature boundaries sliced by month. I'm going to select London to start with.

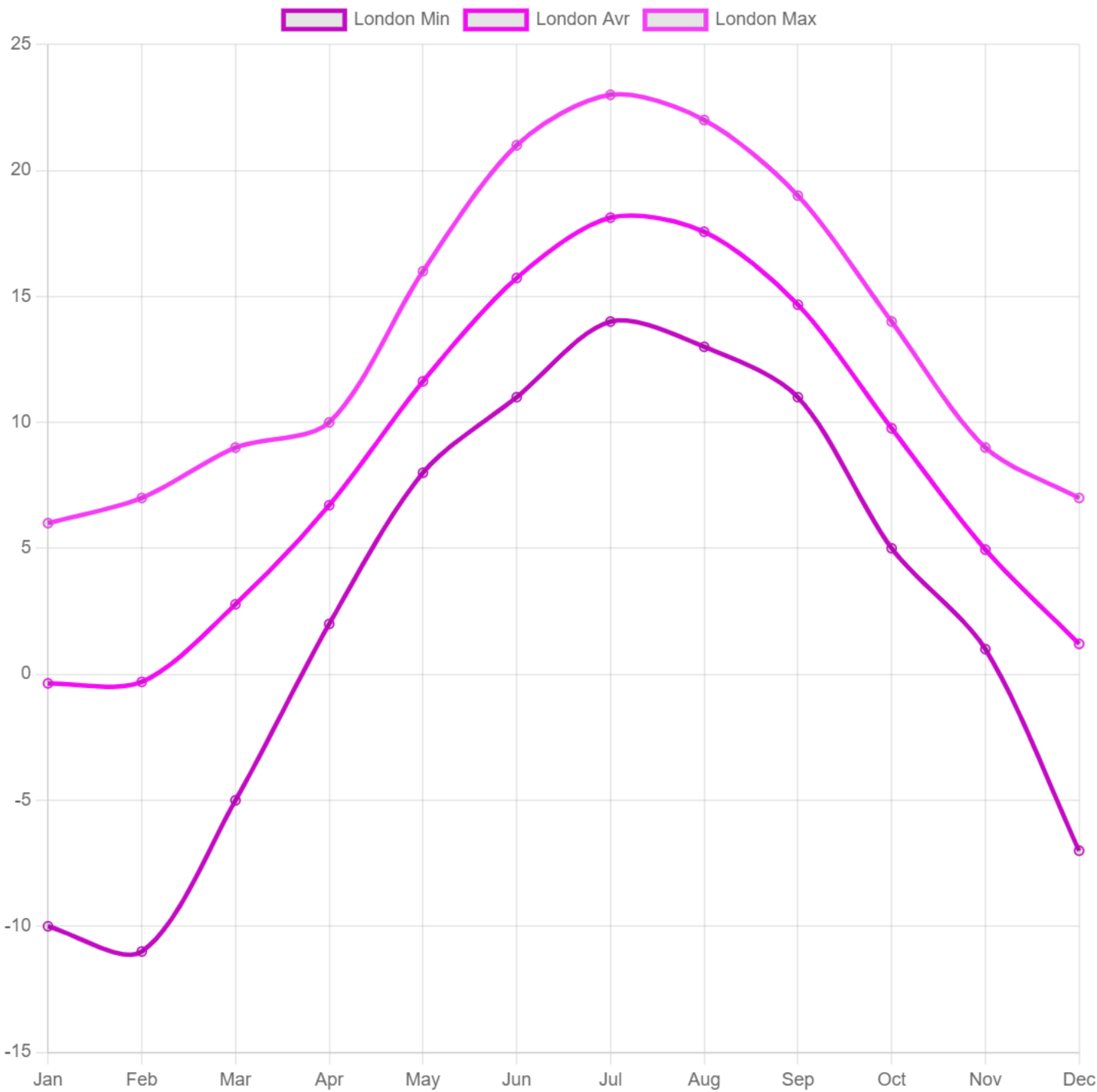
# London Min and Max Temperatures in 18th Century



# London Min, Max & Avr Temperatures in 19th Century

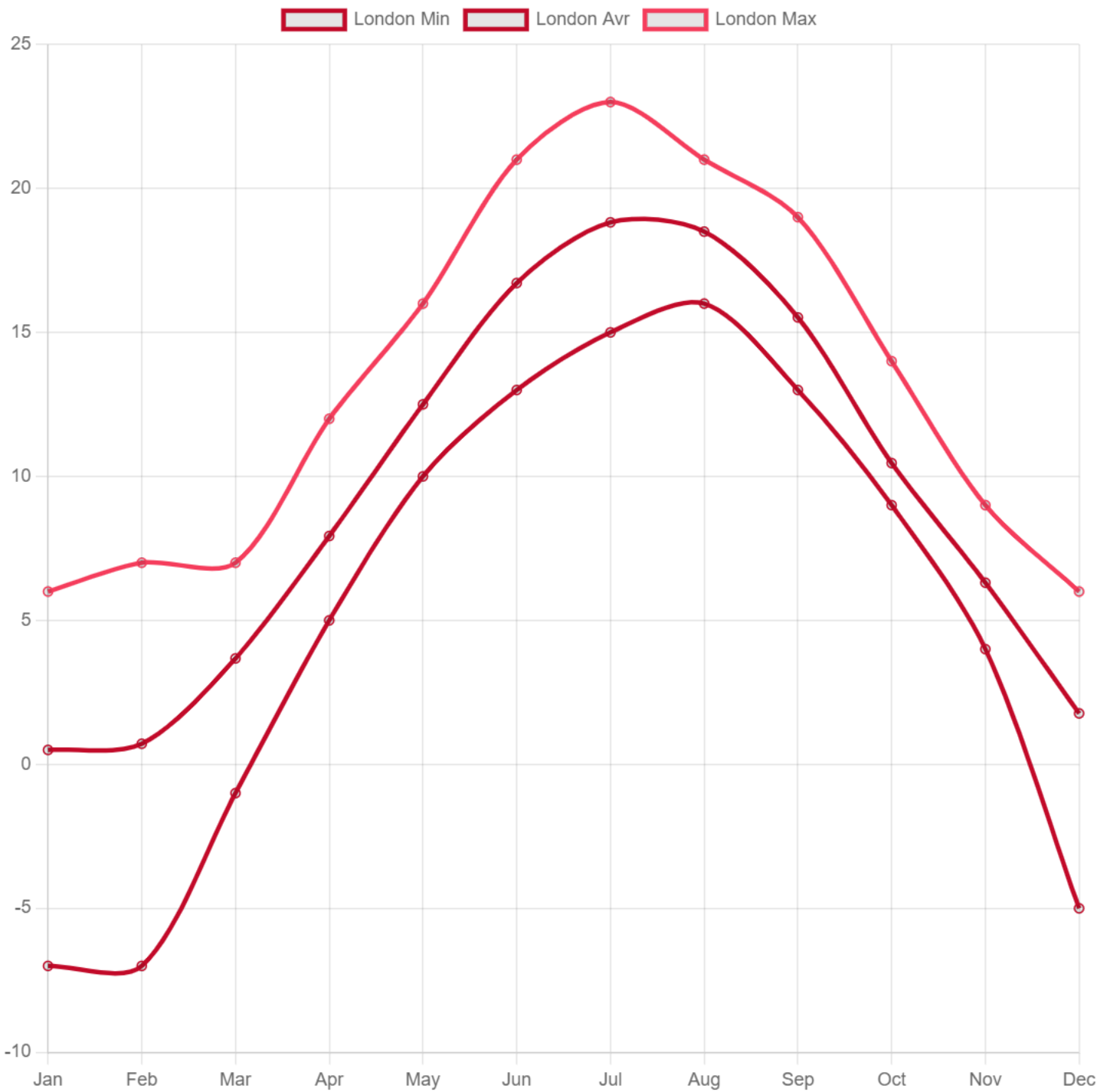


# London Min, Max & Avr Temperatures in 20th Century



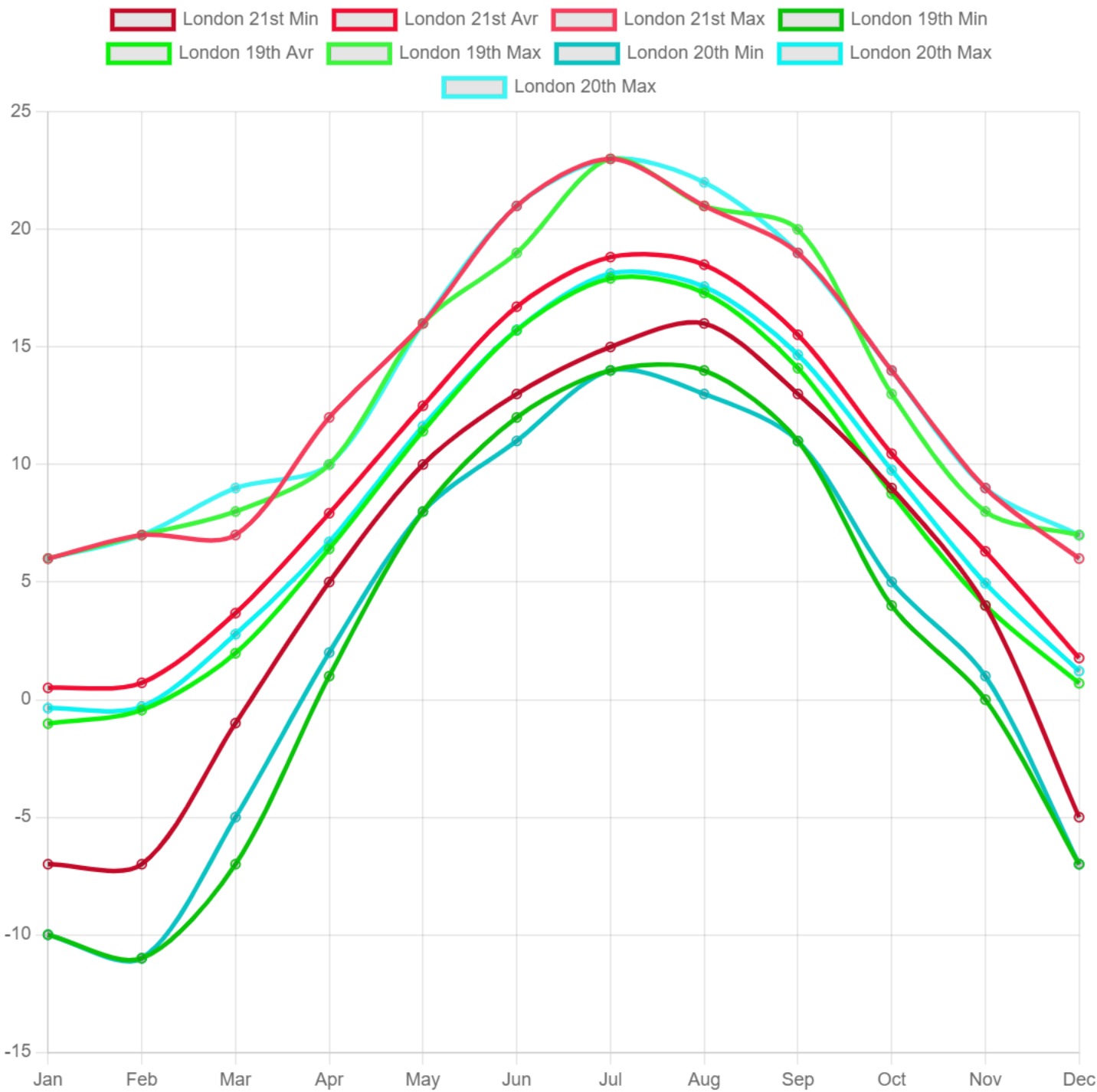


# London Min, Max & Avr Temperatures in 21st Century



# London Min, Max & Avr Temperatures 19th, 20th & 21st Century

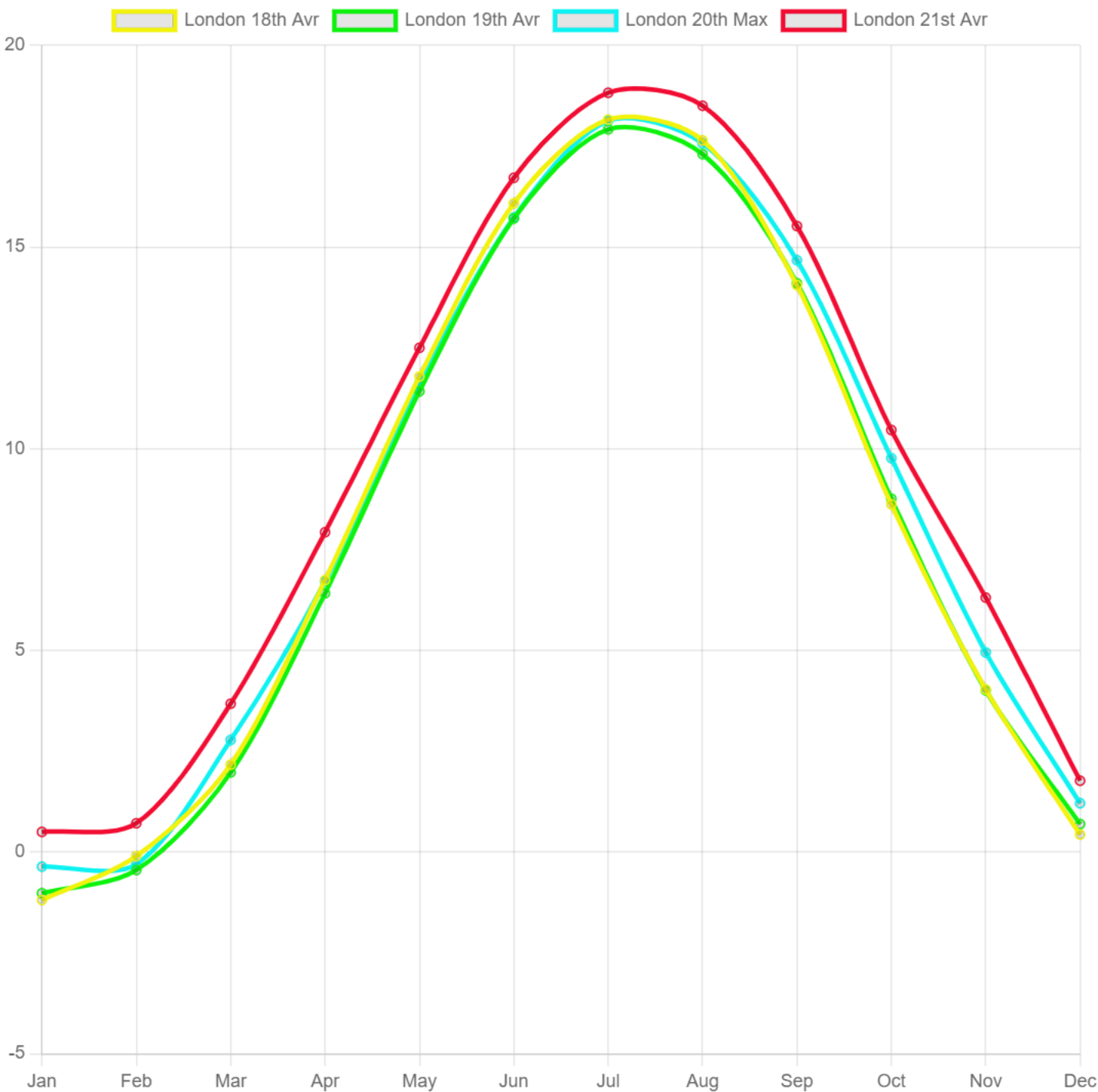
I decided to overlay the decades to see how they compare. It was getting a little noisy so I just included the last three.



# London - Monthly Average Temperatures by Century

It was still a little noisy so I just included the average temperature for each century.

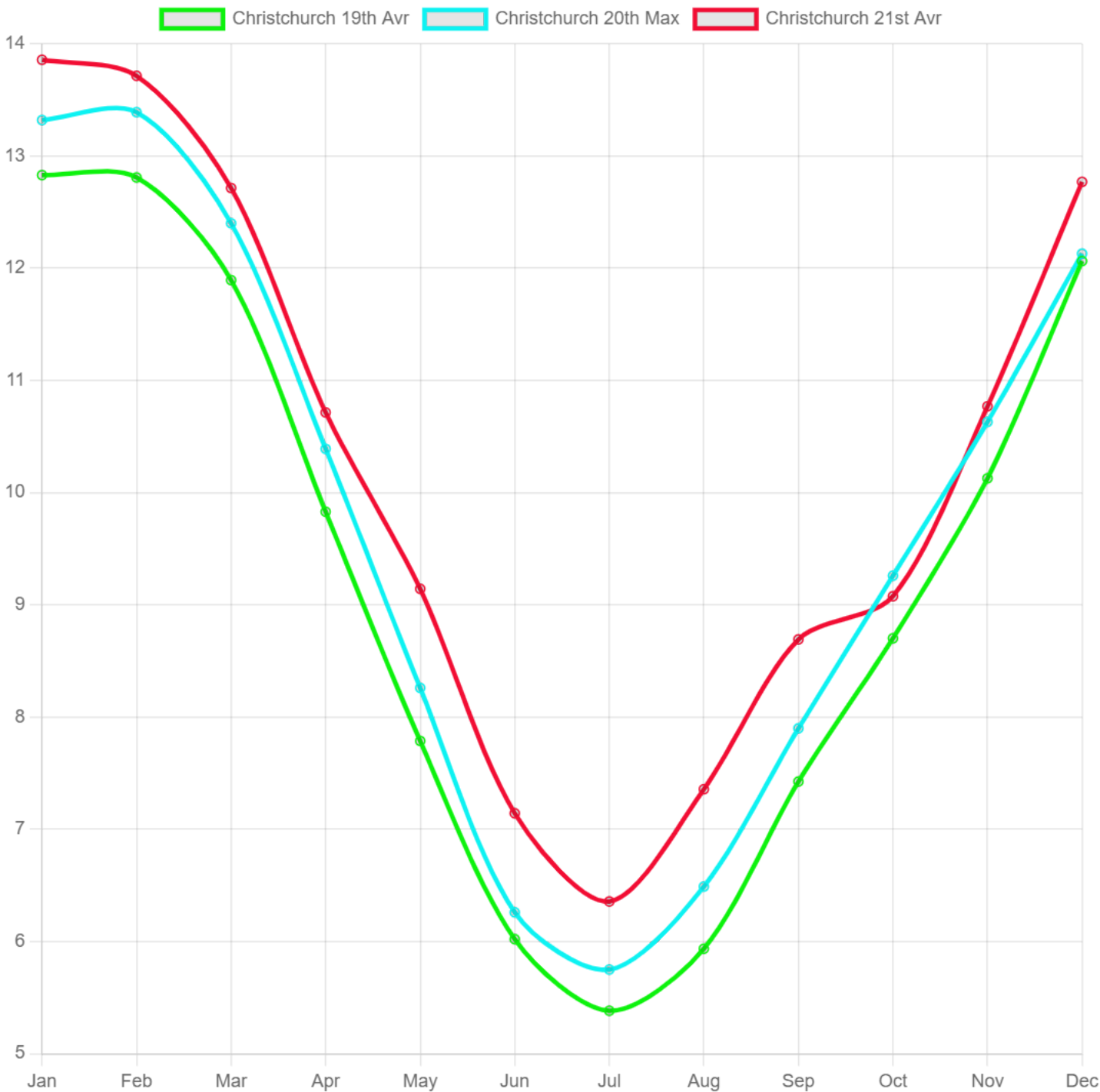
One thing that immediatly stands out is that the 21st century is a step change different to the previous centuries.



# Christchurch - Monthly Average Temperatures by Century

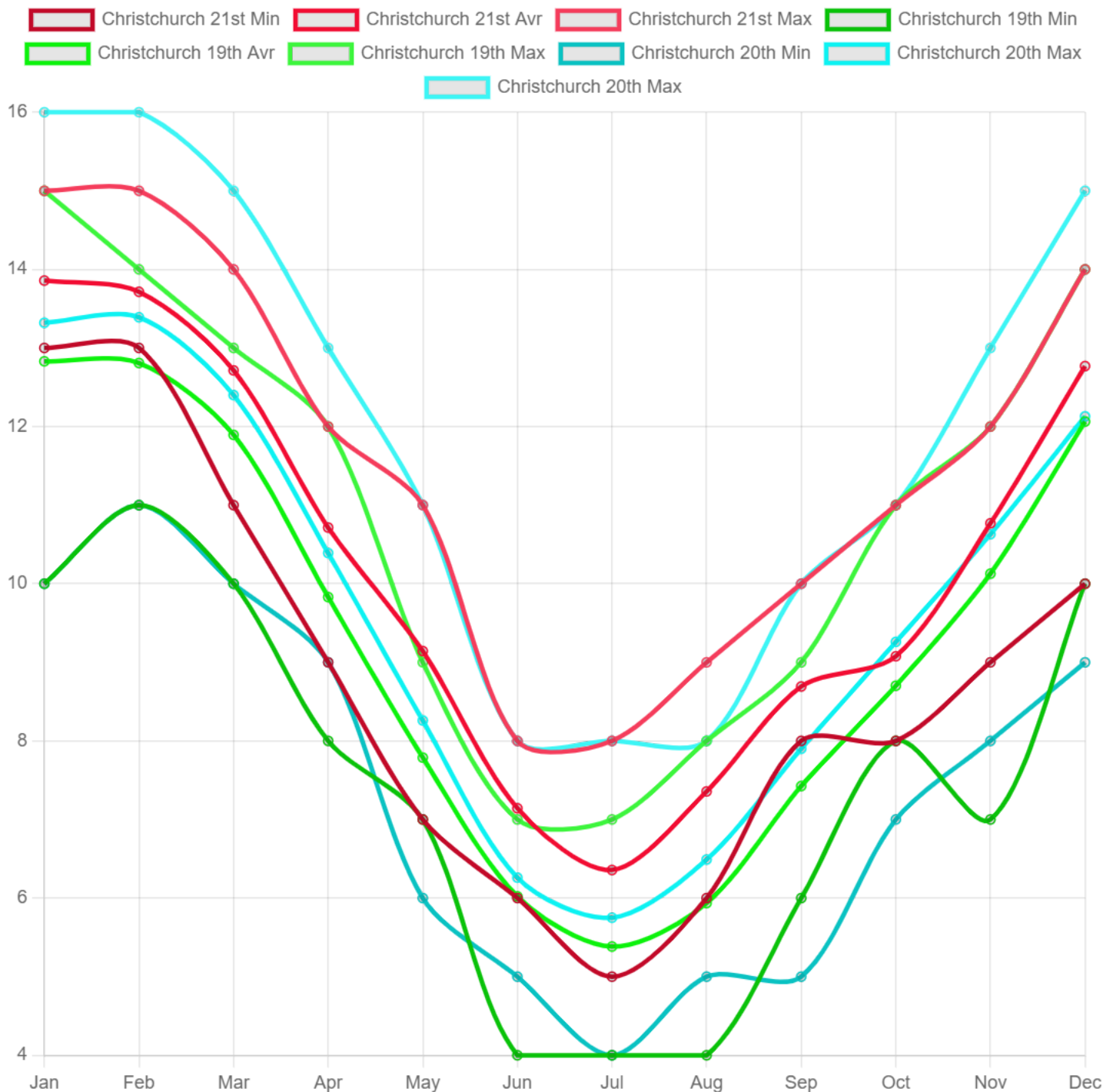
What happens if I select the otherside of the world, Christchurch is probably a good selection for this.

It's interesting seeing the inverse profile to London. The data is lacking 18th century which makes sense. Again, there is a distinctive step change between the centuries in terms of an increase in temperature.



# Christchurch Min, Max & Avr Temperatures centuries

I decided to include the min and max as well. Interestingly the 21st century is higher on average, but the 20th century had higher max values.



It seems like there is a common trend here, but I do want to see more cities first.

In the next part I will make an interactive document that will allow the reader to select country and city from a drop down list and generate a similar set of graphs.