Data summary:
Growth: From Microorganisms to Megacities, Vaclav Smil (2019)

Liam P. Shaw

June 15, 2023

Summary

This document contains graphs fitted from a variety of datasets in *Growth* by Vaclav Smil (2019). Code and data to reproduce analysis are available at `https://github.com/liampshaw/Smil-Growth-2019`. Figure captions and page numbers are taken from the printed version of the book, first edition.

Thanks to Vaclav Smil for sharing these datasets to make this reanalysis possible.

Warning: this reanalysis is my own work. The results will not exactly reproduce the figures or values in Smil (2019). In particular, I am aware of the following issues with some of the fits:

- 1.17.2 - no fit available. I am not aware of an easy off-the-shelf way to fit an asymmetrical logistic in R (not a model I have ever used and not aware of an implementation). I attempted to fit my own specification of model but it was very slow and didn't seem to be converging.

- 4.13 - the fit fails completely. I think since here I am fitting a logistic but should probably try exponential.

- 4.19 - Smil fits logistic but I fit an exponential since the process seems in such early stages of growth, but the fit looks bad.

- 4.28 - ditto the above.

- 5.11 - the logistic fit looks very poor.

1.6.1 US air travel (1930–1980)
Model: Quartic
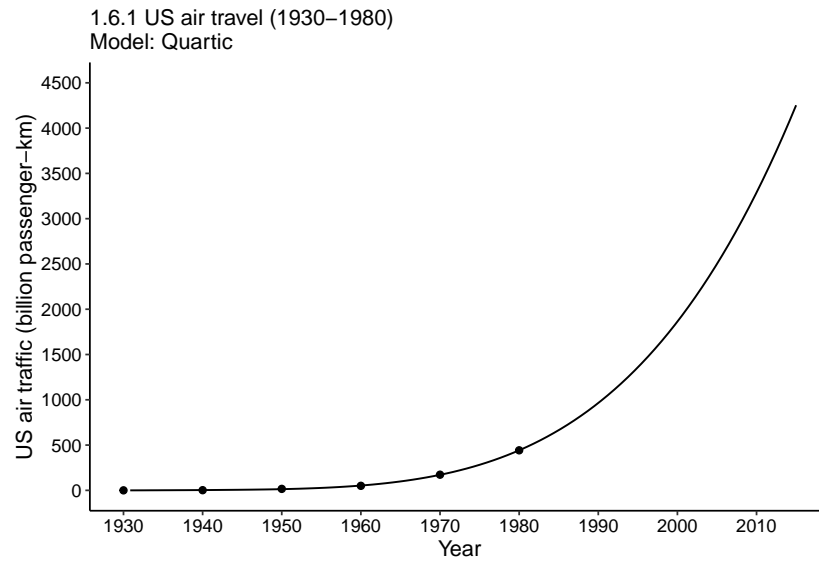
Dataset 1.6.1 (p. 22). Prediction of growth of US air travel (in billions of passenger-kilometers) based on the period 1930-1980. The best fit is a quartic regression. Data from various annual reports by the International Civil Aviation Organization.

1.6.2 US air travel (1930–2015)
Model: Logistic

US air traffic (billion passenger–km)
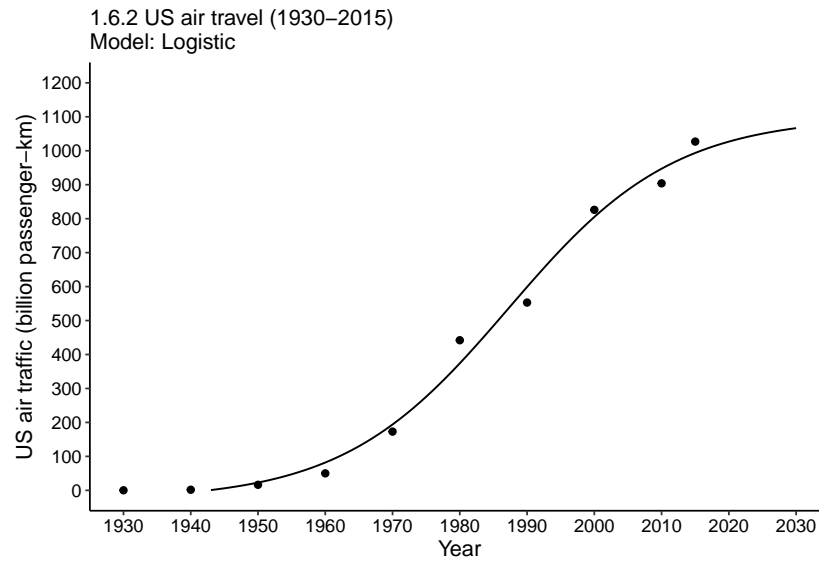
Year

Dataset 1.6.2 (p. 22). Prediction of growth of US air travel (in billions of passenger-kilometers) based on the period 1930-2015. The best fit is a logistic curve with the inflection year in 1987. Data from various annual reports by the International Civil Aviation Organization.

1.13 Growth of a sunflower plant
Model: Logistic
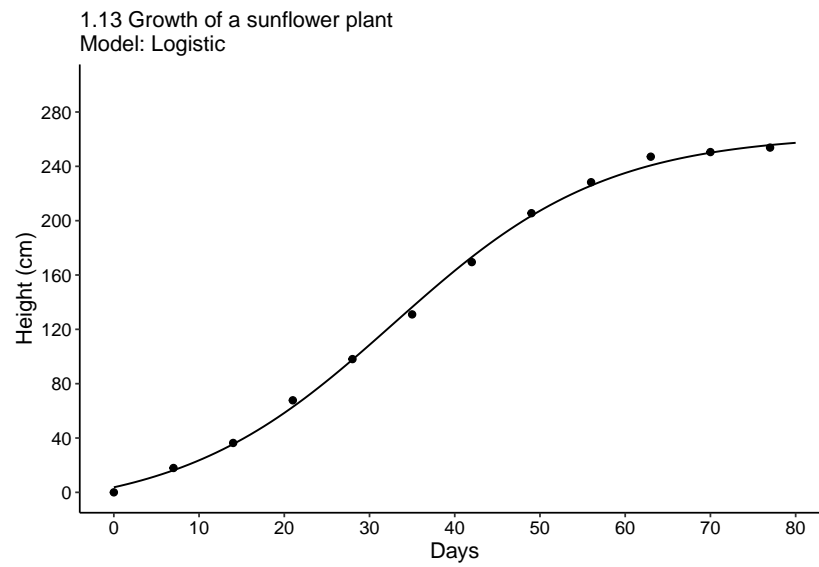
Dataset 1.13 (p. 39). Logistic growth (inflection point at 37.1 days, asymptote at 292.9cm) of a sunflower plant plotted by Reed and Holland (1919).

1.14 US population growth
Model: Logistic
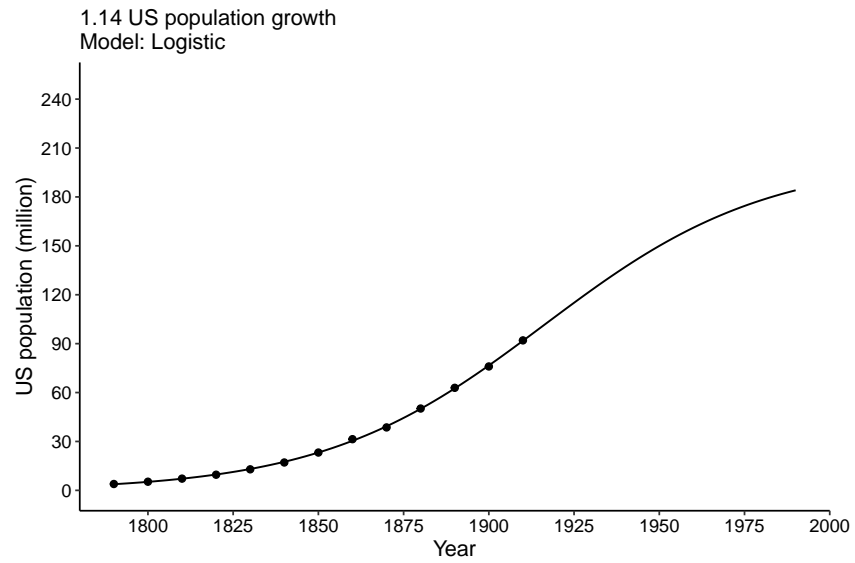
Dataset 1.14 (p. 40). Forecast of US population growth based on the logistic curve (inflection point in 1919, asymptote at 197.3 million) fitted to decennial census data between 1790 and 1930. Data from Pearl and Reed (1920).

1.16 Cruising speed of airliners
Model: Logistic
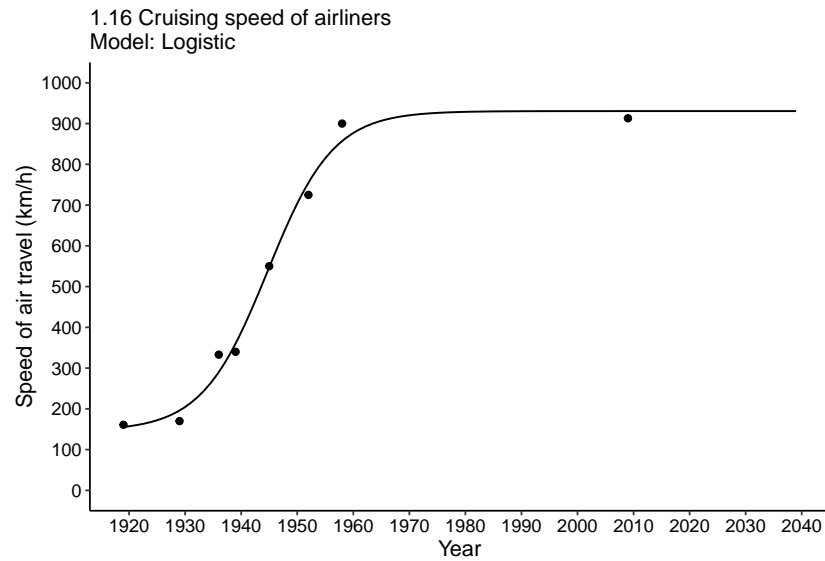
Dataset 1.16 (p. 45). Logistic curve tracing the growth of cruising speed of commercial airliners 1919-2039 (inflection point in 1945, asymptotic cruising speed of 930.8 km/h). Plotted from data on speeds of specific airplanes, starting with KLM's de Havilland DH-16 in 1919 and ending with Boeing 787 in 2009.

1.17.1 Compositions by Mozart (symmetrical logistic)
Model: Logistic
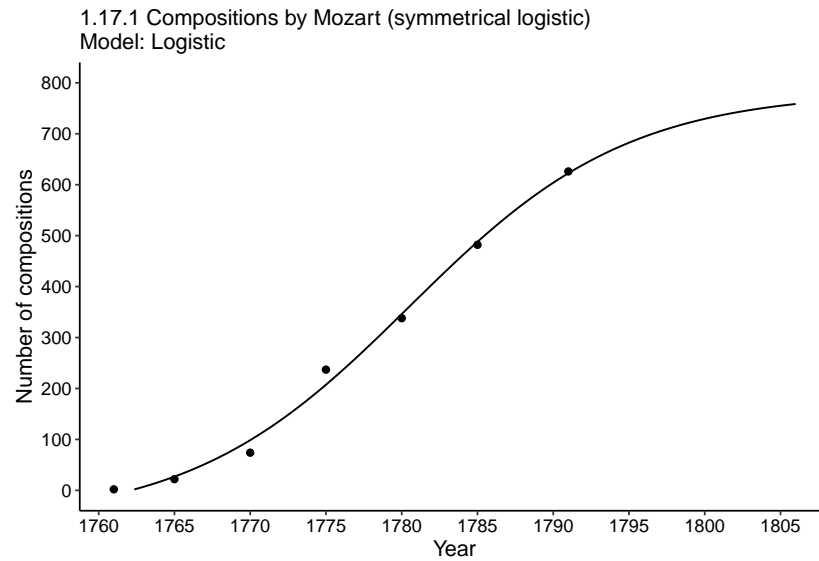
Dataset 1.17.1 (p. 47). Fitting Mozart's oeuvre into growth curves: symmetrical (a) and asymmetrical (b) logistic functions and quadratic (c) and quartic (d) regression all have high degrees of fit ($R^2 = 0.99$) but predict substantially different long-term outcomes for the year 1806 when Mozart (who died in 1791) would have been 50 years old. Compositions by date listed in Giegling et al. (1964).

7

1.17.3 Compositions by Mozart (quadratic)
Model: Quadratic
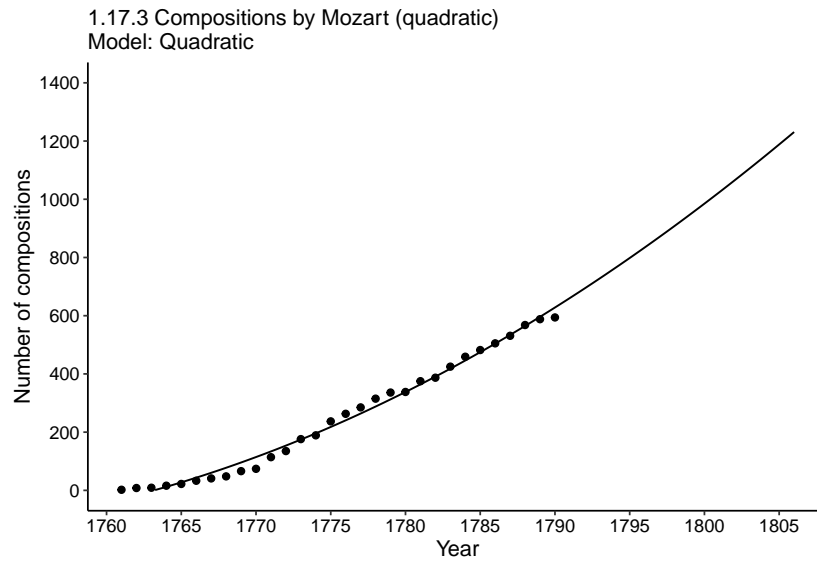
Dataset 1.17.3 (p. 47). Fitting Mozart's oeuvre into growth curves: symmetrical (a)and asymmetrical (b) logistic functions and quadratic (c) and quartic (d) regression all have high degrees of fit ($R^2 = 0.99$) but predict substantially different long-term outcomes for the year 1806 when Mozart (who died in 1791) would have been 50 years old. Compositions by date listed in Giegling et al. (1964).

1.17.4 Compositions by Mozart (quartic)
Model: Quartic
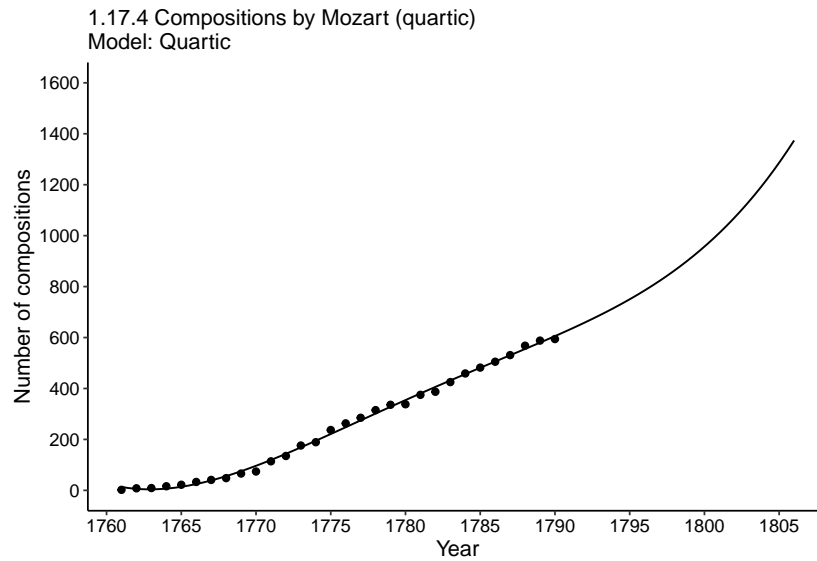
Dataset 1.17.4 (p. 47). Fitting Mozart's oeuvre into growth curves: symmetrical (a)and asymmetrical (b) logistic functions and quadratic (c) and quartic (d) regression all have high degrees of fit ($R^2 = 0.99$) but predict substantially different long-term outcomes for the year 1806 when Mozart (who died in 1791) would have been 50 years old. Compositions by date listed in Giegling et al. (1964).

1.20 Battery energy densities
Model: Logistic
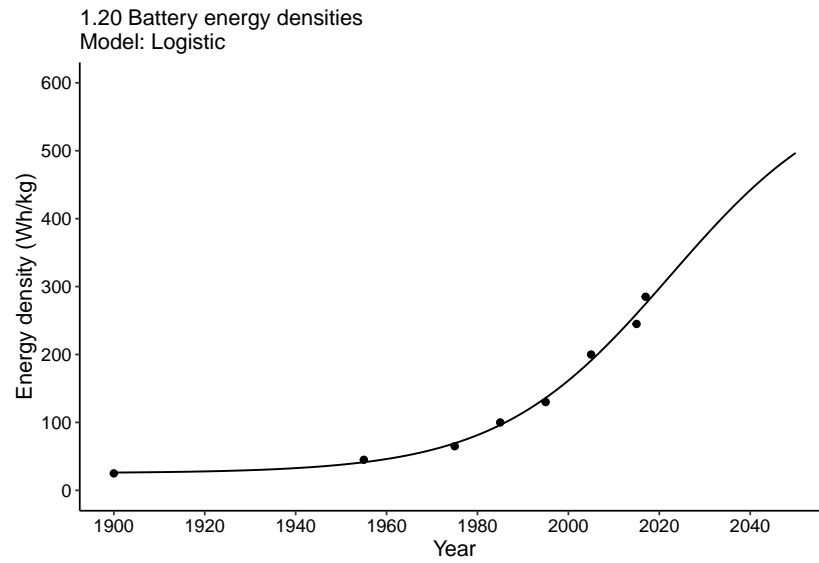
Dataset 1.20 (p. 51). Logistic growth trajectory (inflection point in 2024, asymptote at 625.5 Wh/kg) of battery energy densities, 1900-2017. Plotted from data in Zu and Li (2011) and from subsequent news reports.

2.2 Growth of Escherichia coli O157:H7
Model: Logistic
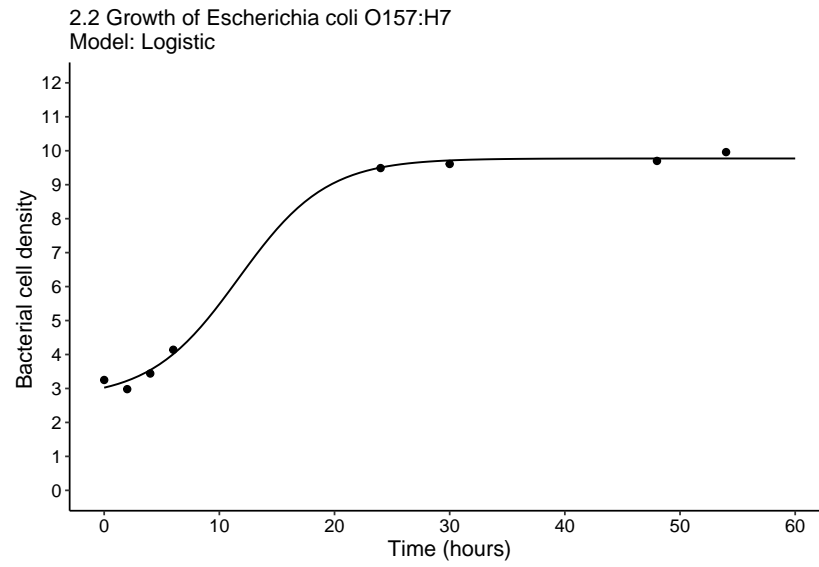
Dataset 2.2 (p. 80). Logistic growth of Escherichia coli O157:H7. Plotted from data in Buchanan et al. (1997).

2.9 Average American wheat yields
Model: Logistic
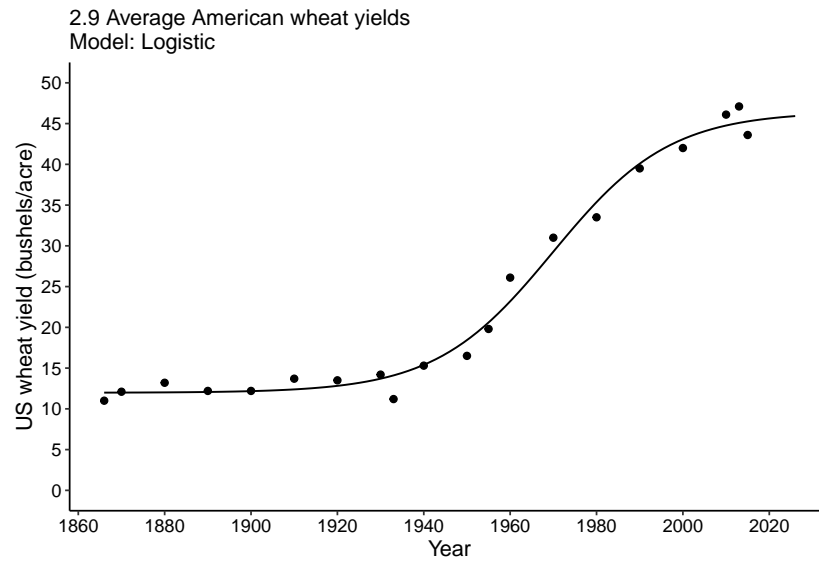
Dataset 2.9 (p. 119). Logistic growth (inflection point in 1970, asymptote at 46.5 bushels/acre) of average American wheat yields, 1866-2015. Data from USDA (2017a).

2.12 Average American grain corn yields
Model: Logistic
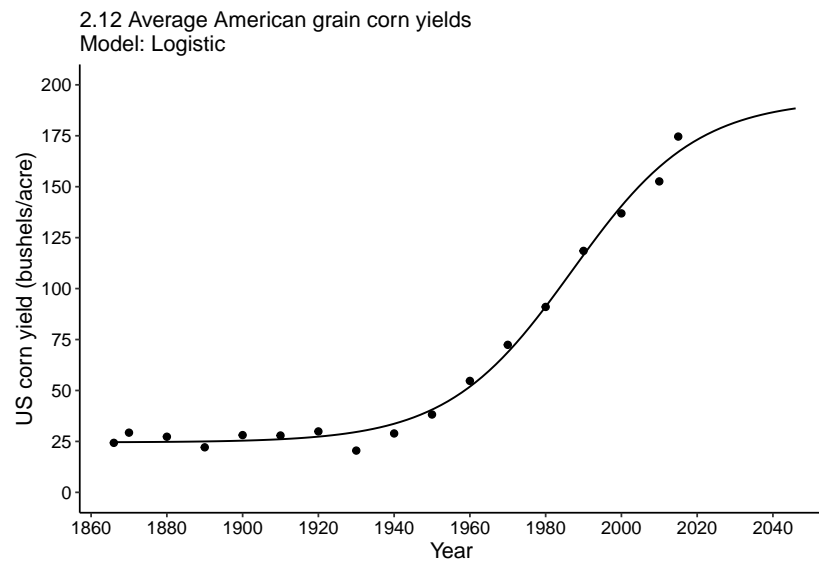
Dataset 2.12 (p. 124). Logistic growth trajectory (inflection point in 1988, asymptote at 194.1 bushels/acre) of average American grain corn yields, 1866-2015. Data from USDA (2017a).

2.19 Average height of 18−year−old Japanese males
Model: Logistic
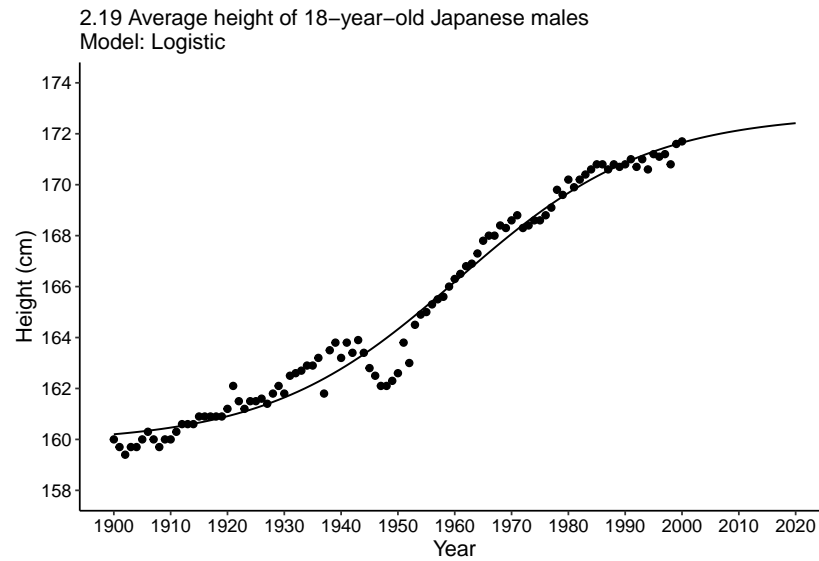
Dataset 2.19 (p. 166). Average height of 18-year-old Japanese males, 1900-2020. Plotted from data in SB (1996). Logistic curve has R^2 of 0.98, inflection point in 1961, and asymptote height is 172.8 cm.

2.20 Obesity prevalence in the US
Model: Logistic
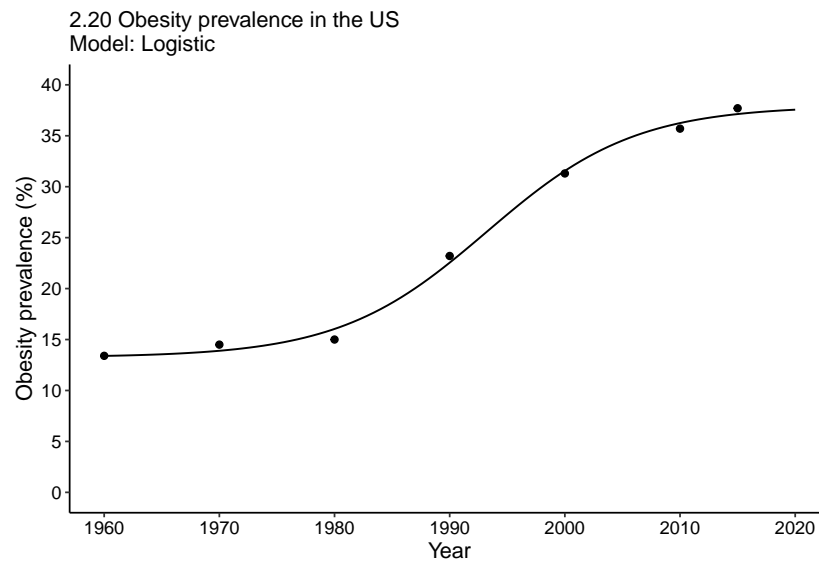
Dataset 2.20 (p. 170). Growing prevalence of obesity in the US. Data from Ogden et al. (2012) and The State of Obesity (2012). The logistic curve had its inflection point in 1993 and its asymptote is at 37.5% of the total population.

3.2 Maximum water turbine capacities
Model: Logistic
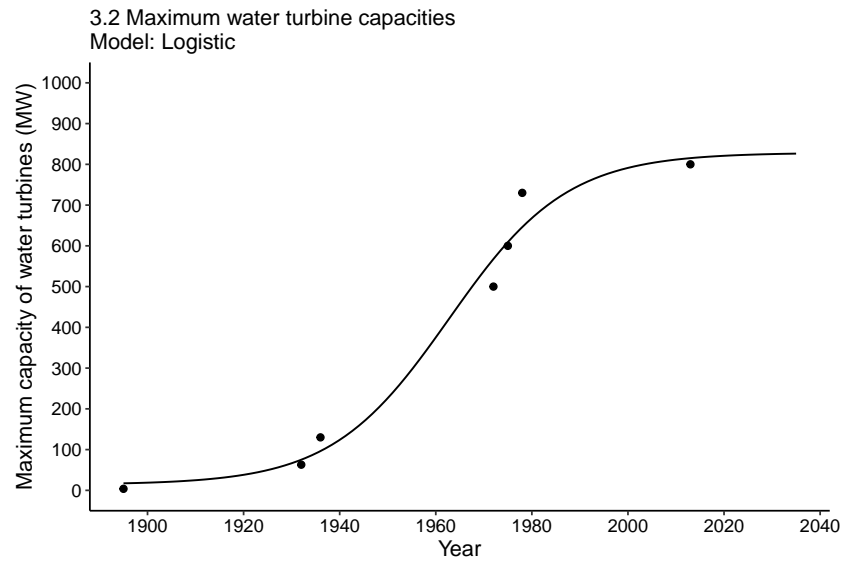
Dataset 3.2 (p. 181). Logistic growth of maximum water turbine capacities since 1895; inflection point was in 1963. Data from Smil (2008) and ICOLD (2017).

16

3.5 Displacement of transatlantic commercial liners
Model: Logistic
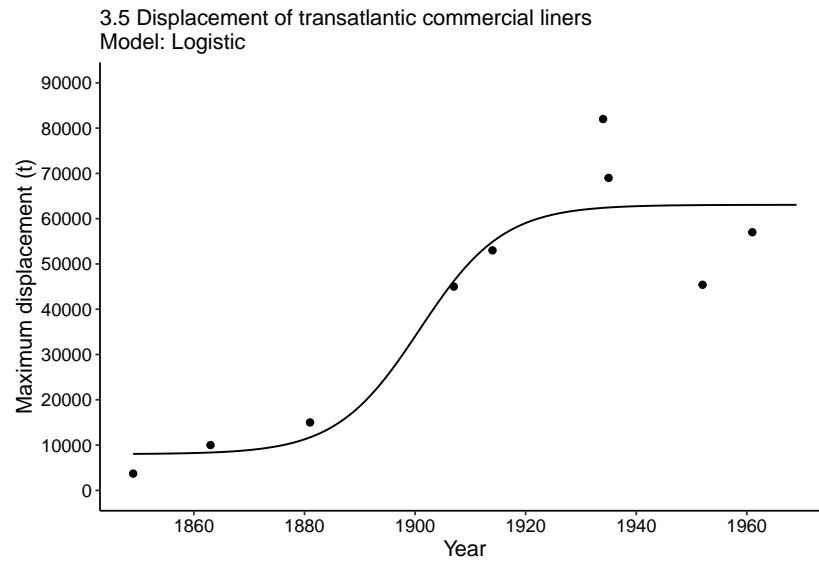
Dataset 3.5 (p. 191). Logistic curve of the maximum displacement of transatlantic commercial liners, 1849-1961 (Smil 2017a).

3.6 Maximum steam turbine capacities
Model: Logistic
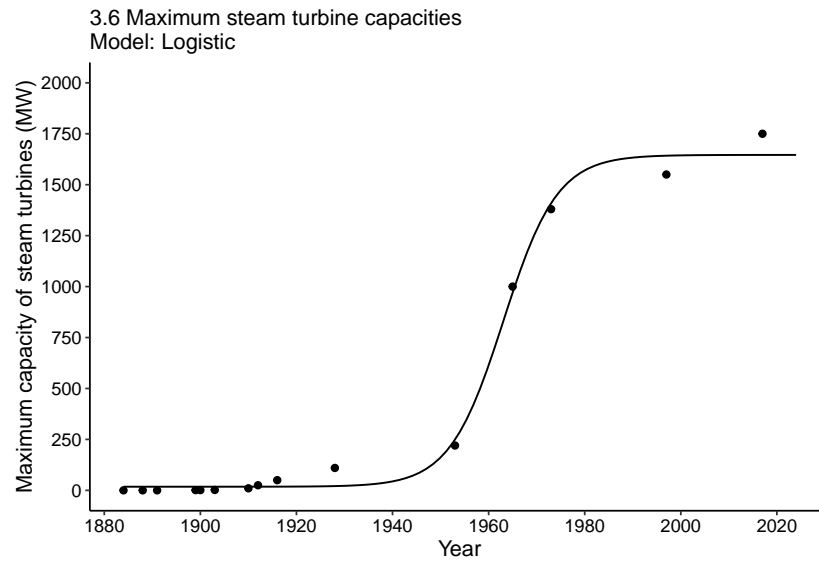
Dataset 3.6 (p. 195). Growth of maximum steam turbine capacities since 1884. Five-parameter [sic] logistic curve, inflection year in 1954, asymptote has been already reached. Data from Smil (2003, 2017a).

3.7 Efficiency of US thermal electricity–generating plants
Model: Logistic
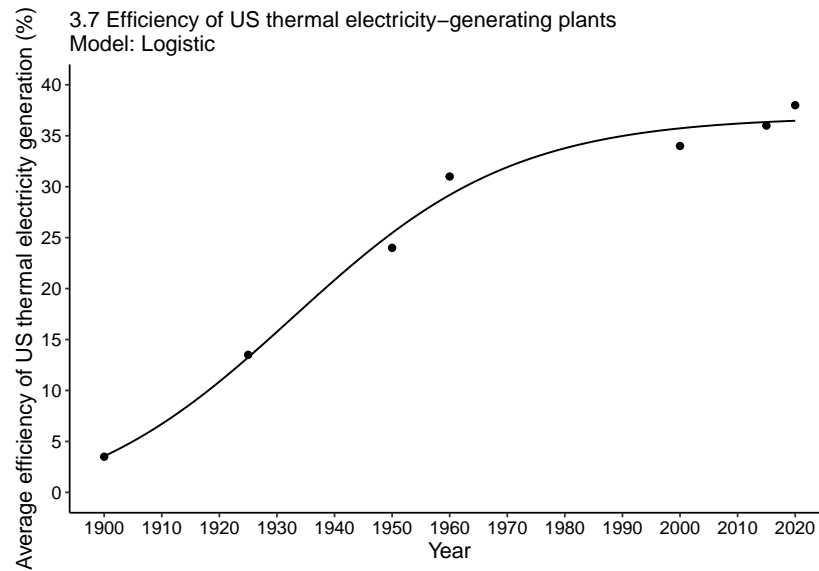
Dataset 3.7 (p. 196). Logistic growth (inflection year in 1933, asymptote at 36.9%) of average efficiency of US thermal electricity-generating plants. Data from Schurr and Netschert (1960) and USEIA (2016).

3.9 Power of US passenger vehicles
Model: Linear
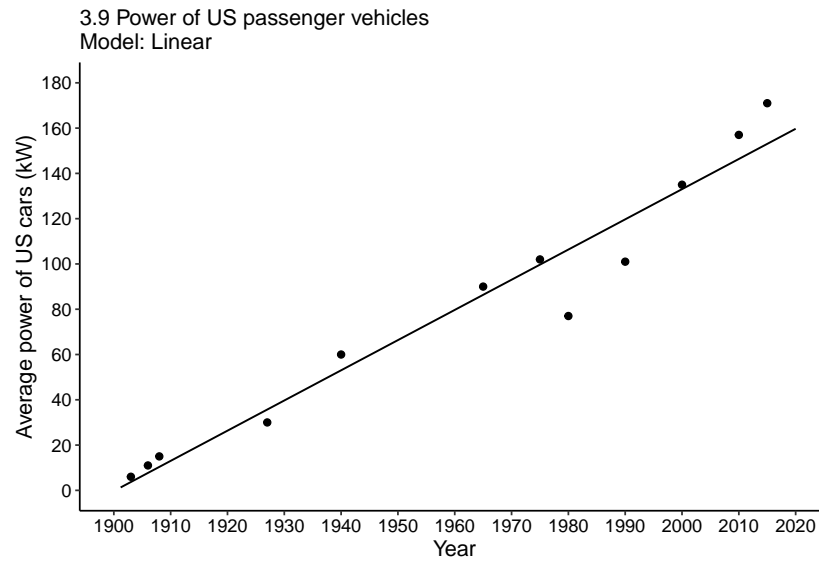
Dataset 3.9 (p. 200). Linear growth of average power of US passenger vehicles, 1903-2020. Data from Smil (2014b) and USEPA (2016b).

3.11 Thrust of jet engines
Model: Linear
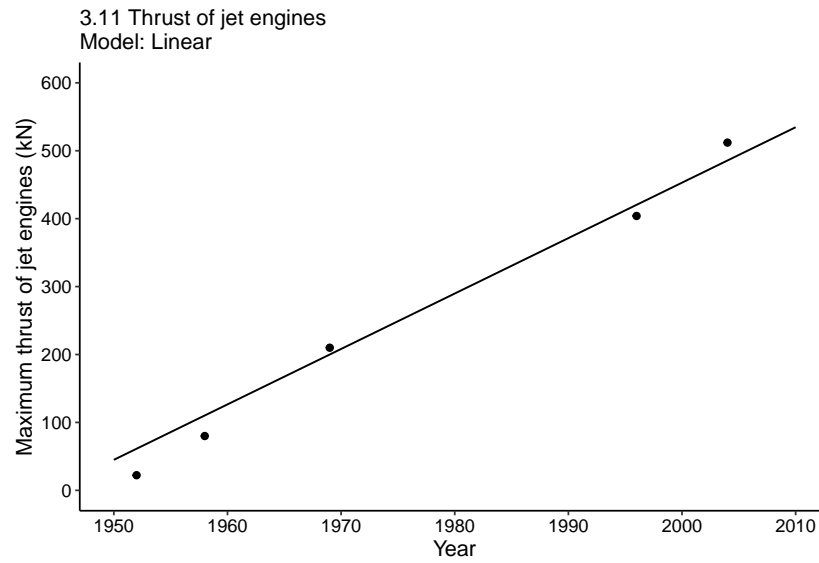
Dataset 3.11 (p. 209). Linear fit of the maximum thrust of jet engines. Data from Smil (2010b).

3.12 Bypass ratio in commercial jetliners
Model: Linear
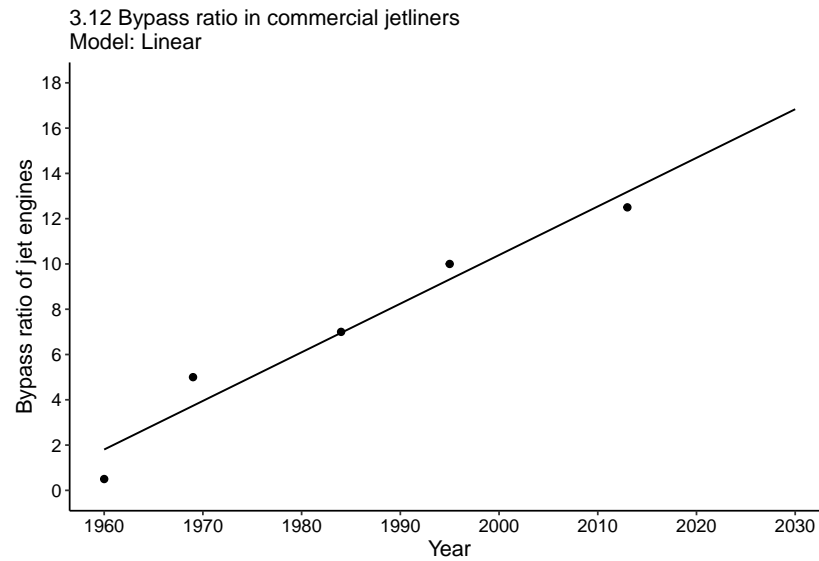
Dataset 3.12 (p. 210). Evolution of the bypass ratio in commercial jetliners. Data from specifications for GE, P&W, and Rolls-Royce engines and from Ballal and Zelina (2003). Maximum ratios have seen linear growth that averaged about 2.2 units per decade.

3.16 Share of power in US manufacturing supplied by electric motors
Model: Logistic
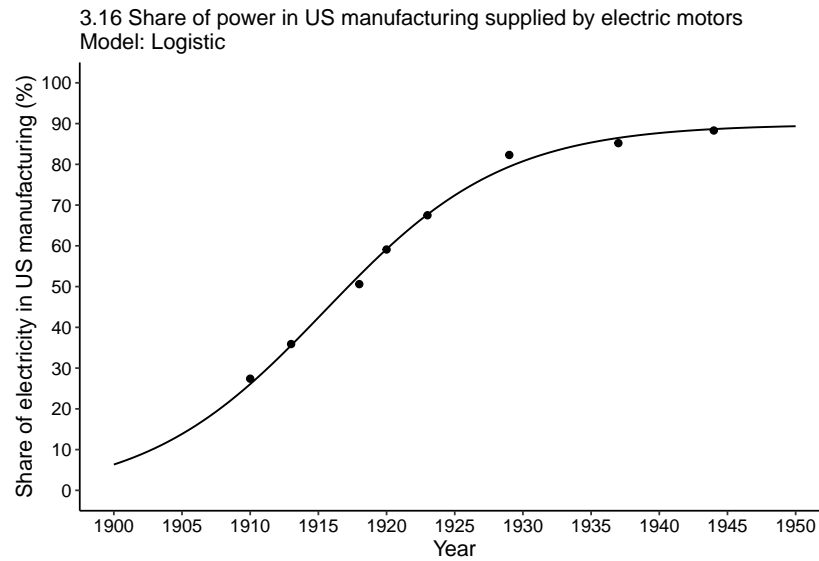
Dataset 3.16 (p. 221). Logistic fit (inflection point in 1916, asymptote of 89.9%) of the share of power in US manufacturing supplied by electric motors, 1909-1950. Data from Daugherty (1927) and Schurr et al. (1990).

4.6.1 Skyscraper heights
Model: Exponential
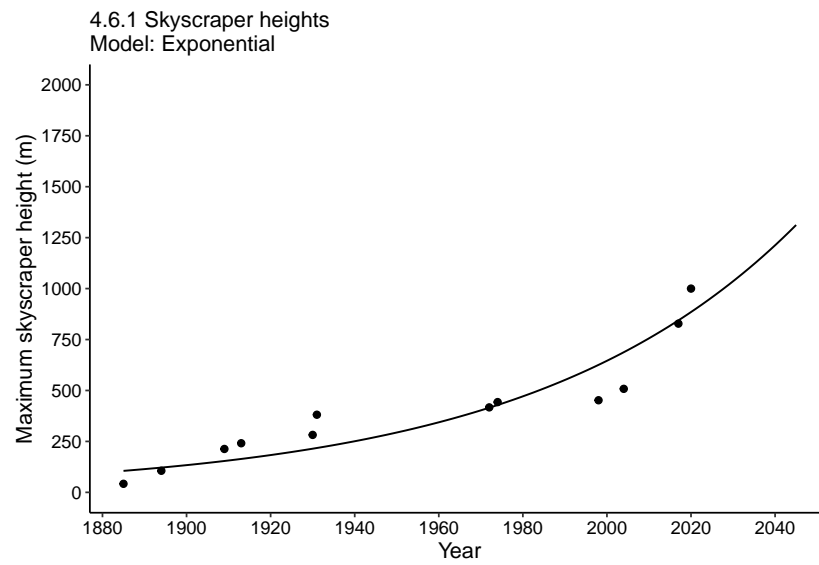
Dataset 4.6.1 (p. 245). Logistic curve of the growth of maximum skyscraper height. Data from Landau and Condit (1996) and Skyscraper Center (2017).

4.6.2 Skyscraper heights
Model: Quartic
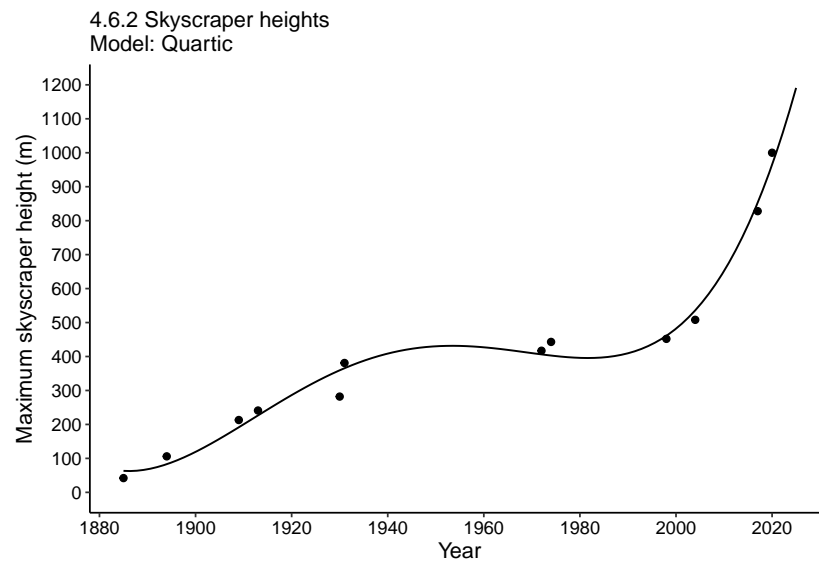
Dataset 4.6.2 (p. 245). Polynomial regression of the growth of maximum skyscraper height. Data from Landau and Condit (1996) and Skyscraper Center (2017).

25

4.7 Buildings taller than 200m
Model: Exponential
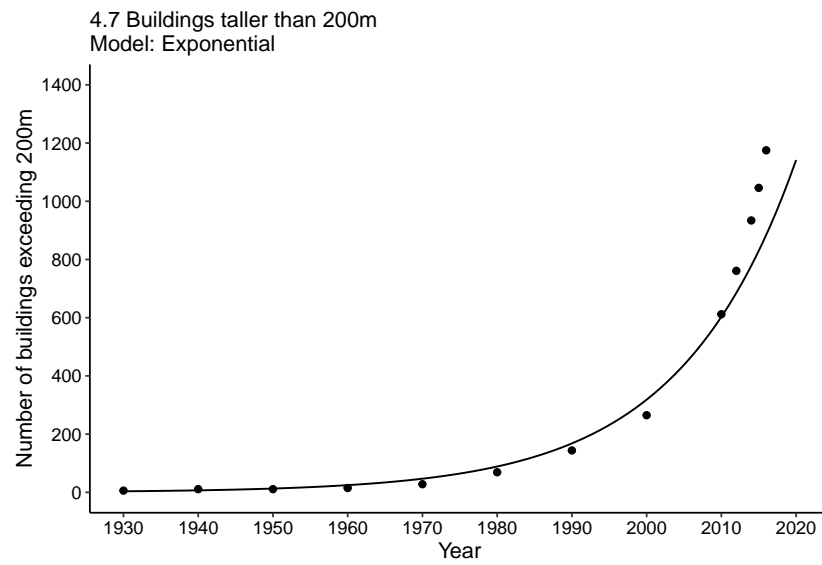
Dataset 4.7 (p. 247). Growth of the total number of buildings taller than 200m. Logistic curve in its early stage of ascent. Data from Emporis (2017).

4.8 Area of American houses
Model: Logistic
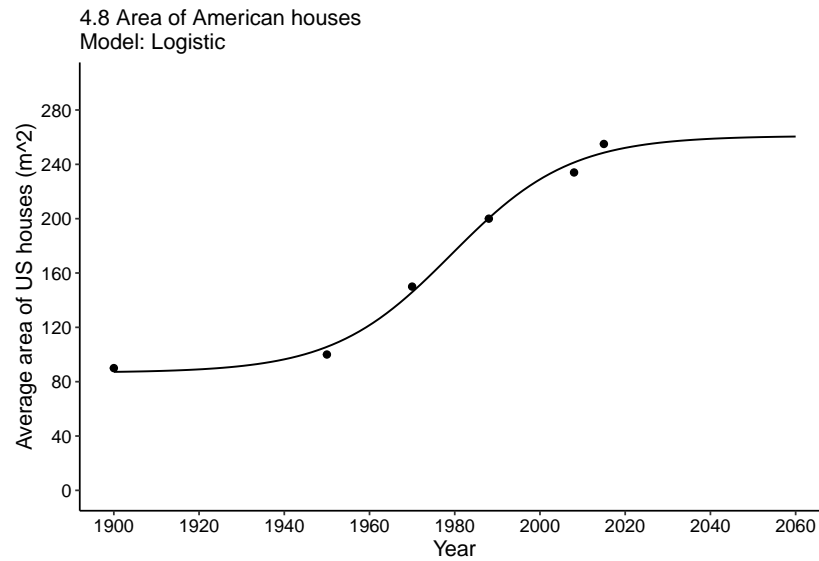
Dataset 4.8 (p. 250). Growth of the average area of American houses since 1900. Logistic curve had the inflection year in 1979 and its asymptote is about 260 m^2. Data from Wilson and Boehland (2005) and USCB (2016a).

4.9 Longest railway tunnels
Model: Logistic
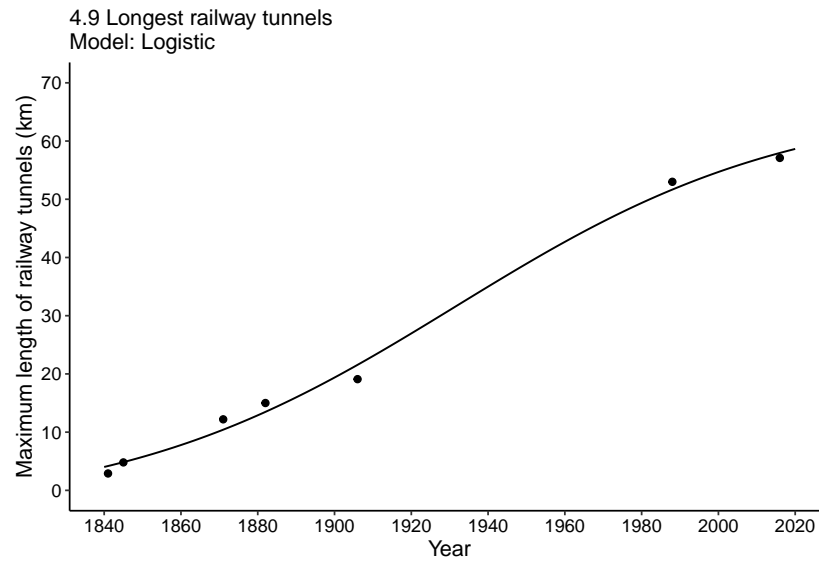
Dataset 4.9 (p. 256). Growth of the longest railway tunnels, 1840-2020. Data mostly from Beaver (1972) and Onoda (2015).

4.10 Longest suspension bridges
Model: Logistic
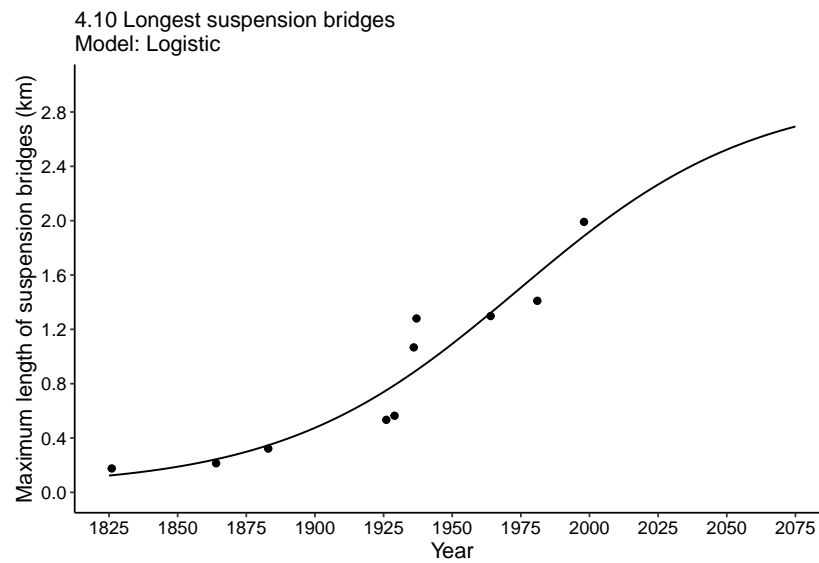
Dataset 4.10 (p. 258). Growth of the longest suspension bridges since 1825. Data mostly from History of Bridges (2017).

4.11 German autobahns
Model: Logistic
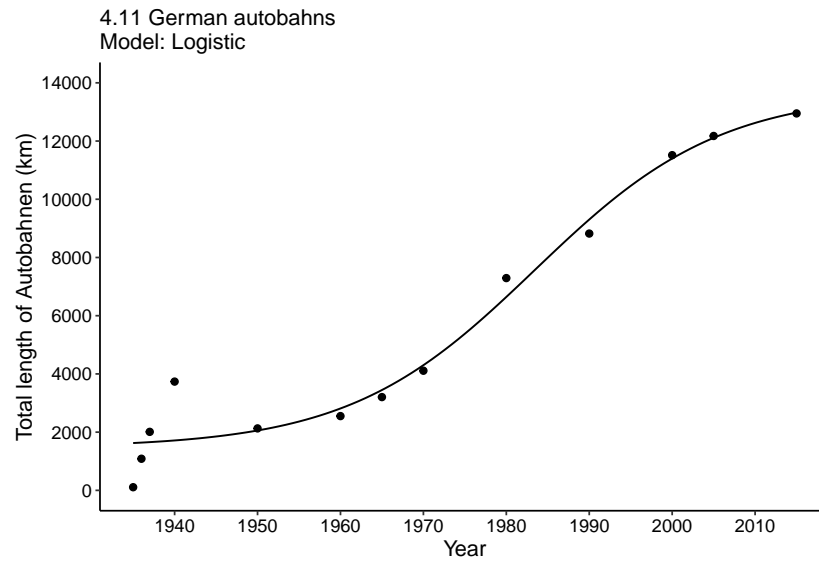
Dataset 4.11 (p. 260). Growth curve of German Autobahnen, 1935-2015. Data from Zeller (2007) and Bundesamt für Statistik.

4.12 Paved US roads
Model: Logistic
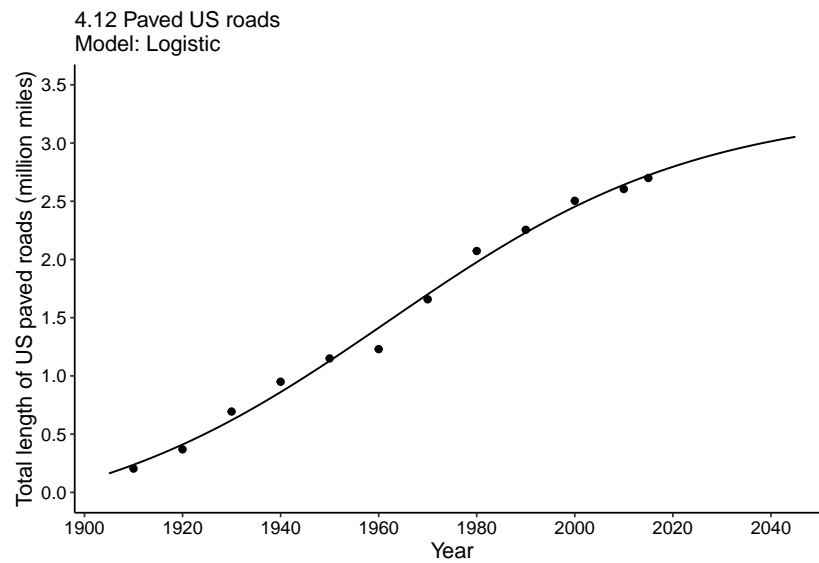
Dataset 4.12 (p. 261). Growth of the total length of paved US roads since 1905. Plotted from data in USBC (1975) and subsequent volumes of US Statistical Abstract.

4.13 Highways in China
Model: Logistic
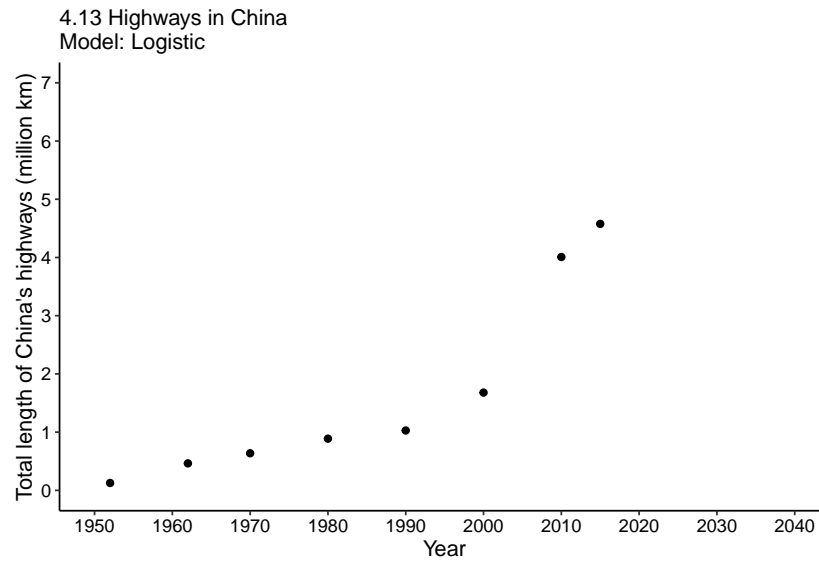
Dataset 4.13 (p. 262). Growth of the total length of highways in China. Logistic curve with the inflection point in 2007 and asymptote about 30% above the 2015 total. Data from NBS (2000, 2016).
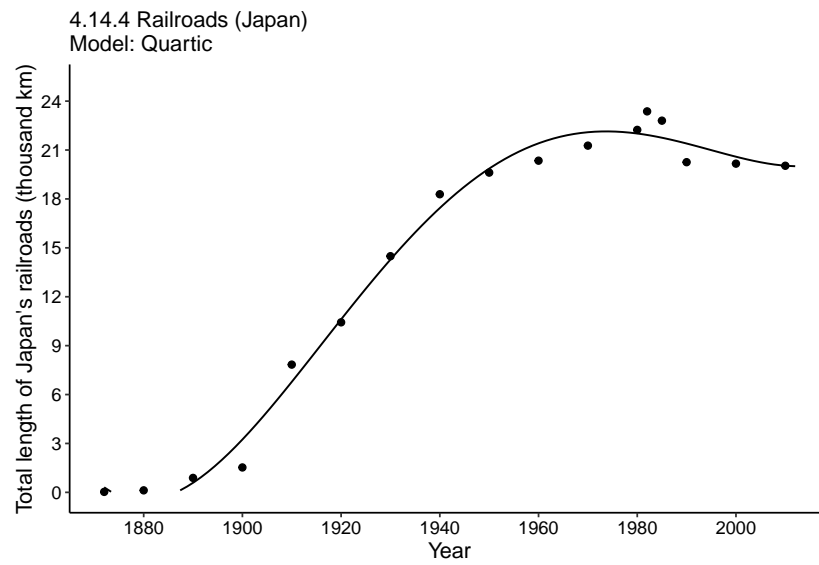
4.14.1 Railroads (US)
Model: Quartic
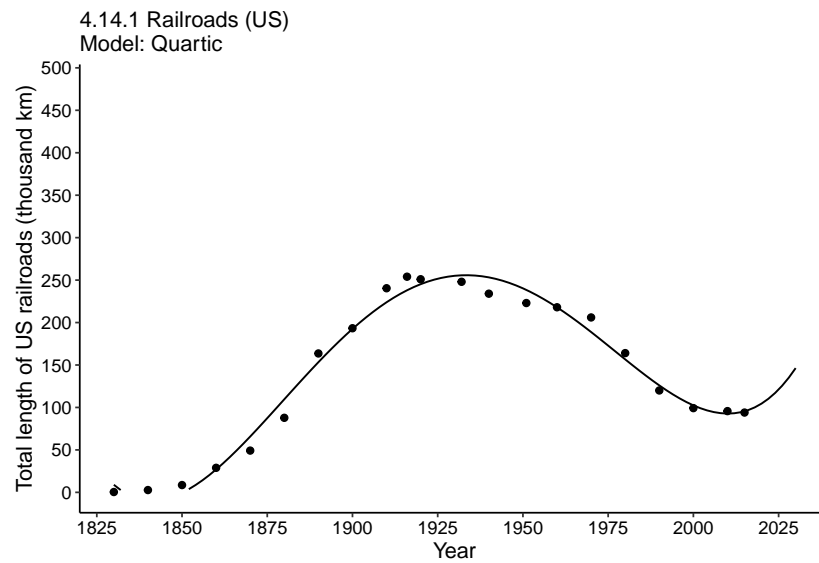
Dataset 4.14.1 (p. 263). Growth curves of the total length of railroads in the US. Plotted from data in Mitchell (1998).
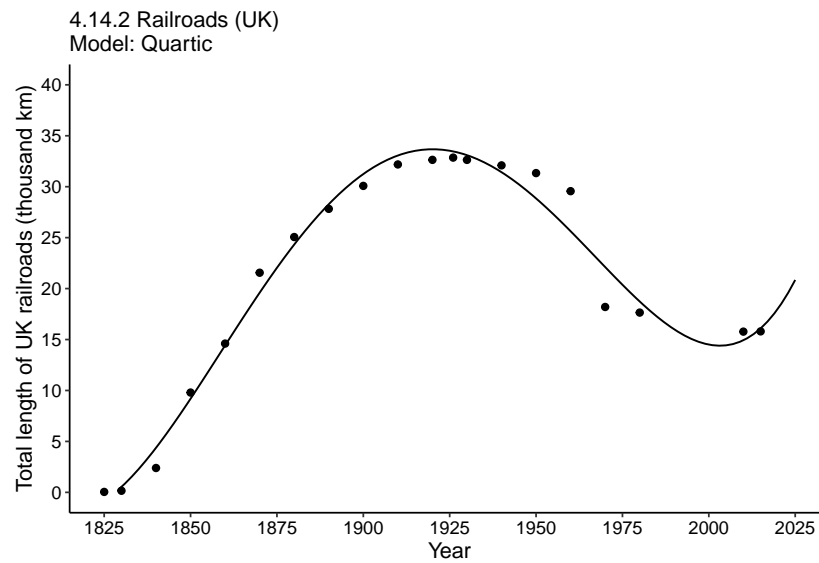
Dataset 4.14.2 (p. 263). Growth curves of the total length of railroads in the UK. Plotted from data in Mitchell (1998).

4.14.3 Railroads (France)
Model: Quartic
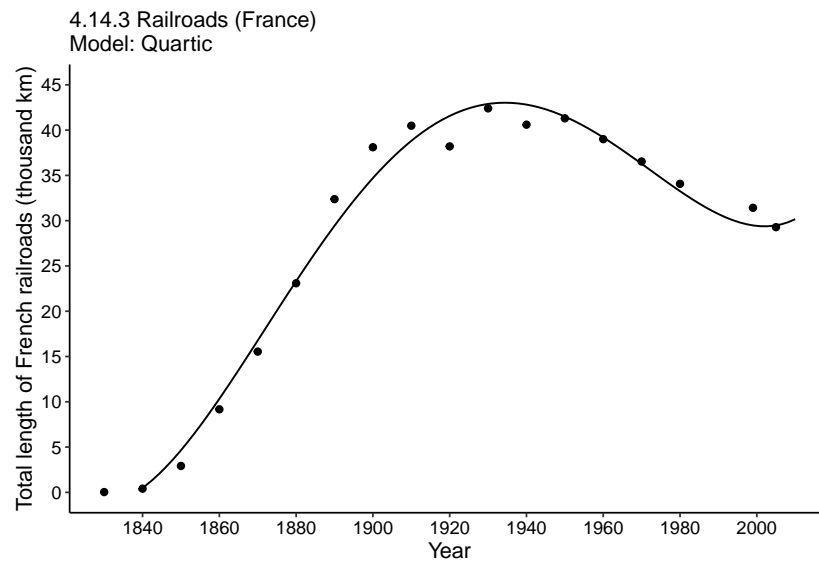
Dataset 4.14.3 (p. 263). Growth curves of the total length of railroads in France. Plotted from data in Mitchell (1998).

4.14.4 Railroads (Japan)
Model: Quartic

Dataset 4.14.4 (p. 263). Growth curves of the total length of railroads in Japan. Plotted from data in Mitchell (1998).

4.17 Passengers on Japan's trains
Model: Logistic
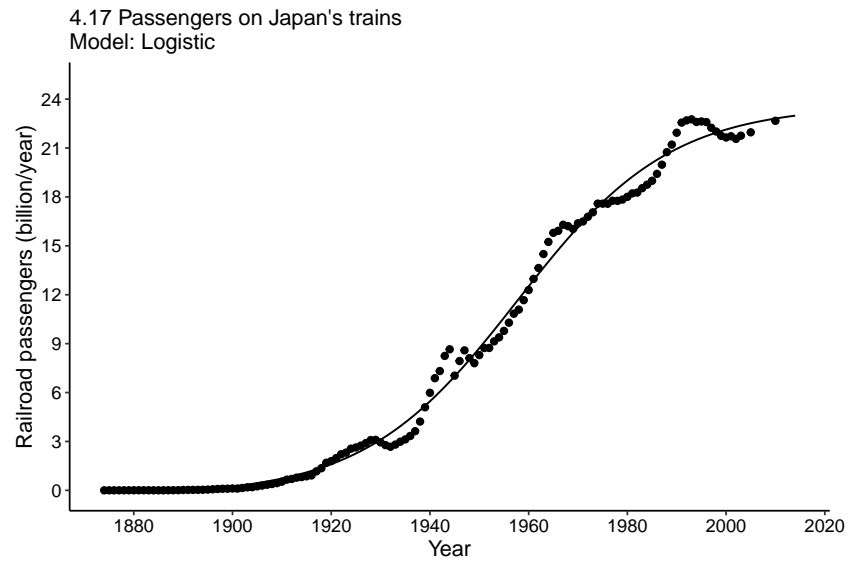
Dataset 4.17 (p. 278). Passengers on Japan's trains, 1874-2014. Plotted from data available at SB (1996, 2017a). Logistic curve with the inflection year in 1958 and with the 2015 total only a fraction of 1 percent below the asymptotic level.

4.18 Passenger capacity of commercial airplans
Model: Logistic
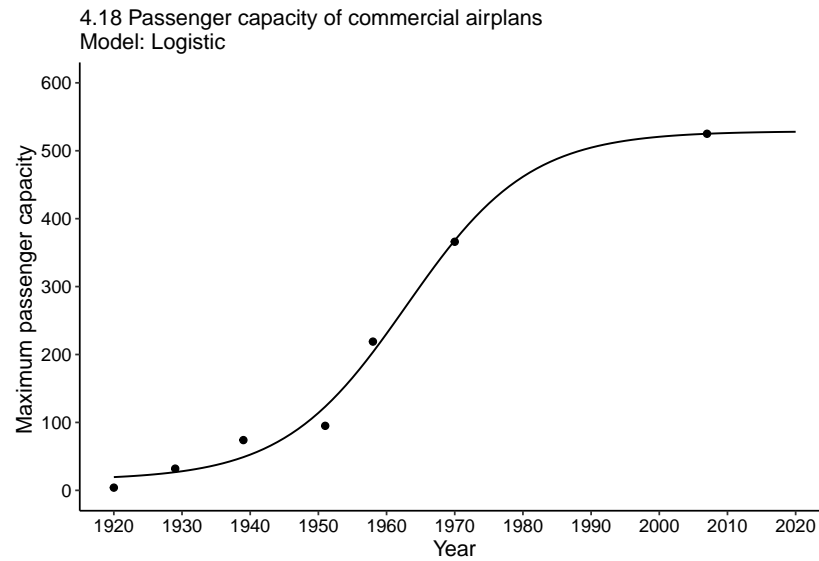
Dataset 4.18 (p. 282). Logistic fit of the maximum passenger capacity of commercial airplans: from KLM's de Havilland DH.16 (four passengers) in 1920 to Airbus 380 (544 passengers in three classes, 868 maximum) in 2007. Plotted from individual airplane specifications.

4.19 Global civil aviation traffic
Model: Exponential
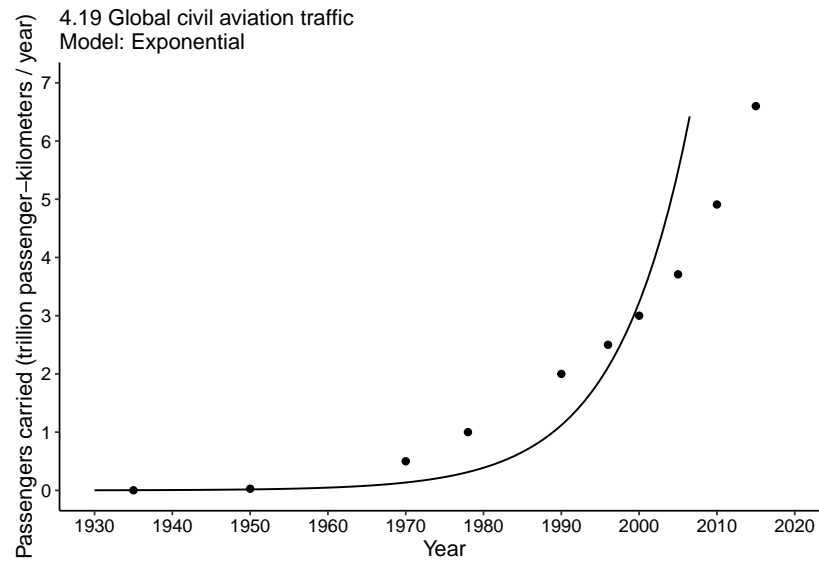
Dataset 4.19 (p. 283). Growth of global civil aviation traffic (domestic and inter-national flights) measured in terms of passenger-kilometers. Logistic curve in its early stage indicates further substantial growth in the decades ahead. Data from ICAO (2016) and from earlier annual reports.

4.24 Mobile phone sales
Model: Logistic
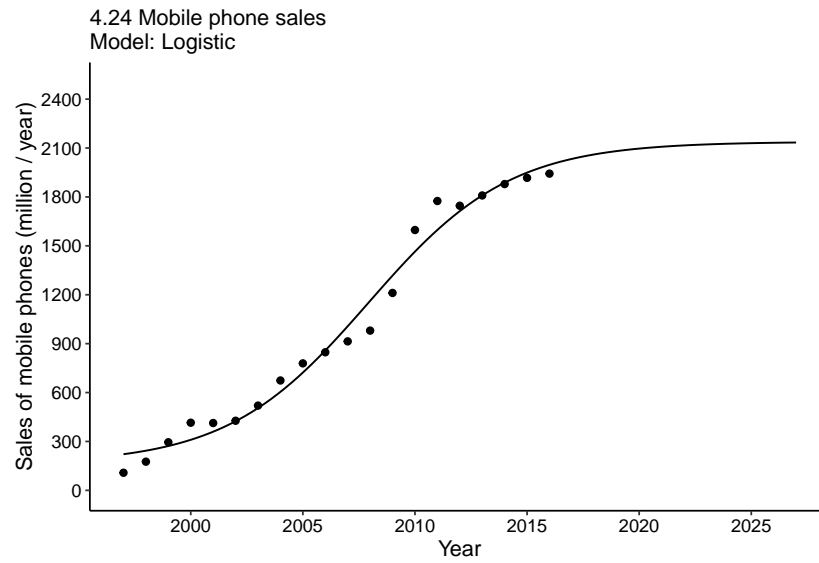
Dataset 4.24 (p. 296). Growth of annual sales of all mobile phones since 1997. The trajectory fits a logistic curve that inflected in 2008 and that is now approaching its asymptotic value. Data from GSMArena (2017).

4.25 Smartphone sales
Model: Logistic
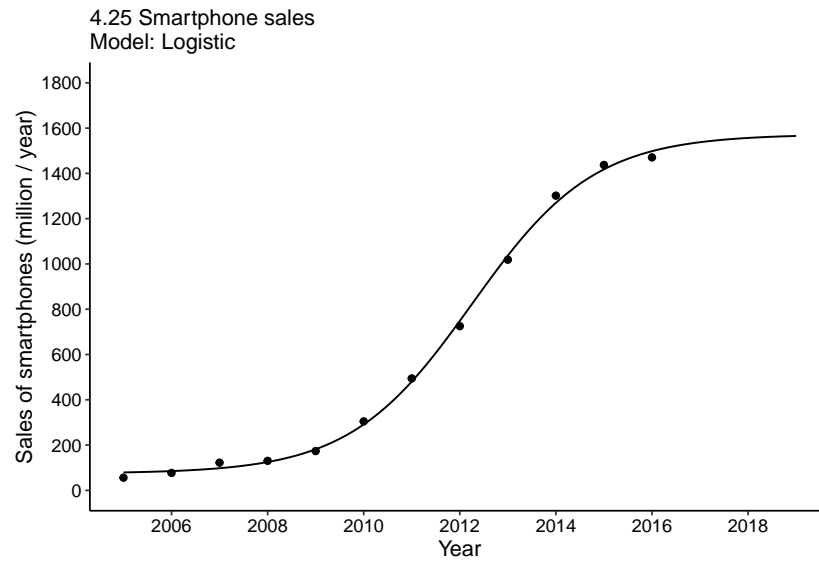
Dataset 4.25 (p. 297). The growth of annual sales of smartphones since the year 2005 has followed perfectly a logistic curve with the inflection year in 2012 and with the asymptote less than 10% above 2016 sales. Data from Canalys (2007) and Meeker (2017).

4.26 Internet hosts
Model: Logistic
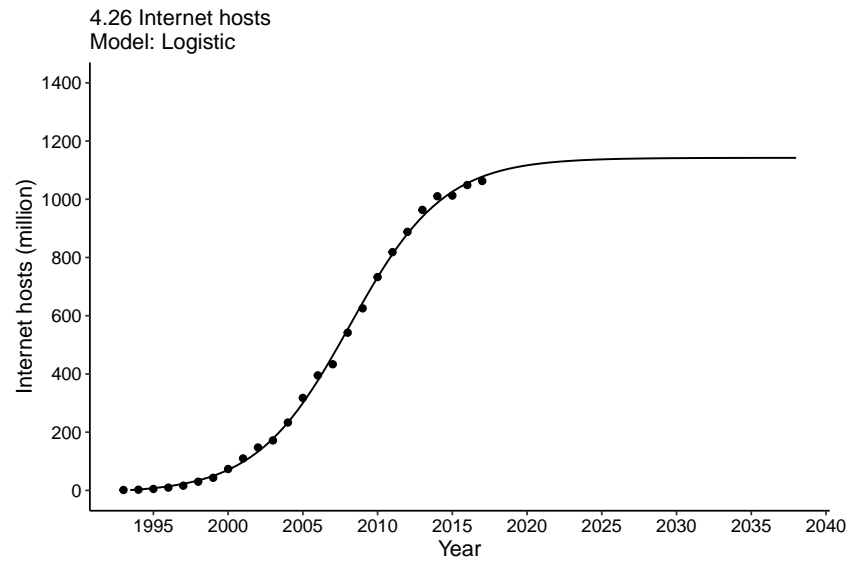
Dataset 4.26 (p. 299). The post-1993 growth of Internet hosts has followed a logistic curve with the inflection year in 2008 and with the asymptotic value less than 10% above the 2017 total. Data from ISC (2017).

4.27 Internet hosts
Model: Gaussian
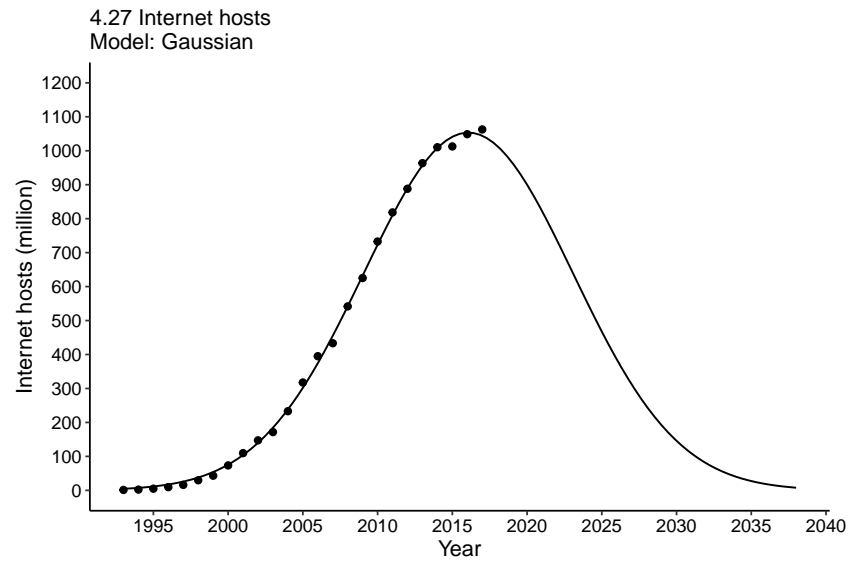
Dataset 4.27 (p. 299). The growth of Internet hosts also fits a Gaussian curve peaking in 2016 and returning to negligible values before 2040. Such a development seems quite unlikely - unless a new mode of hosting takes over. Data from ISC (2017).

4.28 Internet traffic
Model: Exponential
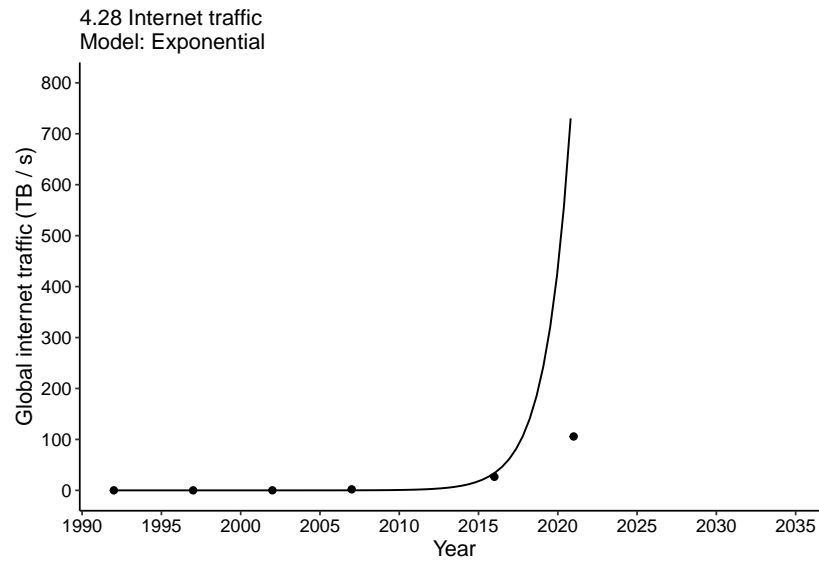
Dataset 4.28 (p. 300). Post-1992 growth of Internet traffic in TB/s fits a logistic curve in its early stages of growth indicating further substantial gains in decades ahead. Data from CISCO (2017).

5.5 Global population growth 1700–2500
Model: Logistic
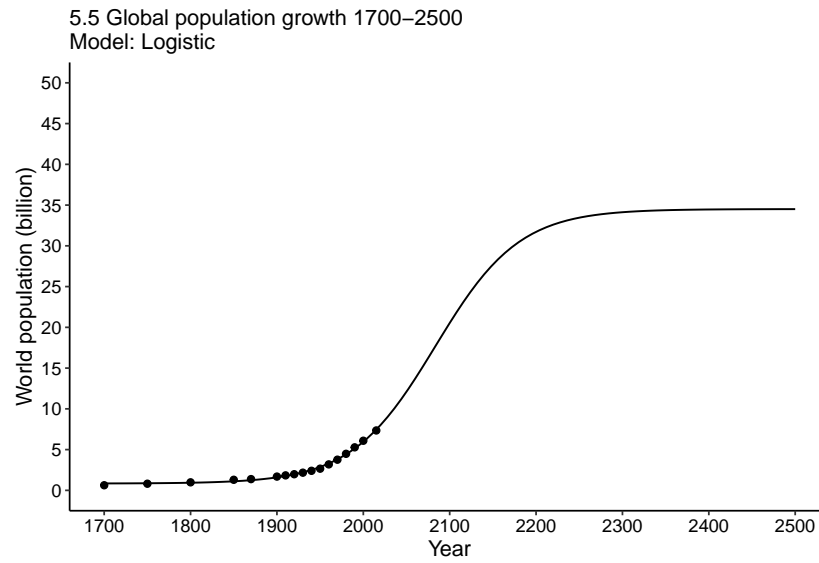
Dataset 5.5 (p. 332). Global population growth, 1700-2500. Logistic curve with a distant inflection point (in 2105) and asymptote of 45.2 billion.

5.6 Global share of urban population 1800–2050
Model: Logistic
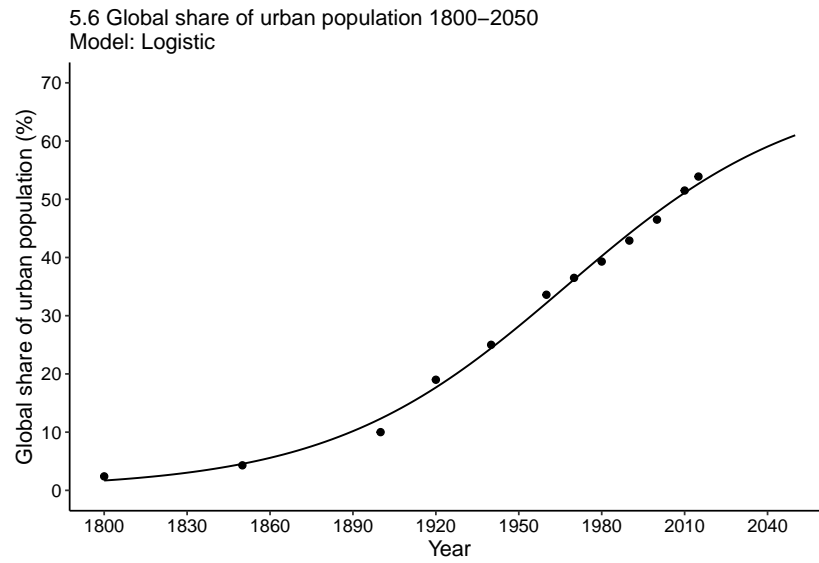
Dataset 5.6 (p. 337). Global share of the urban population, 1800-2050. Plotted from various UN statistics. Logistic curve with the inflection point in 1969 and asymptote of 70.9%.

5.7 Urban population in the US since 1790
Model: Logistic
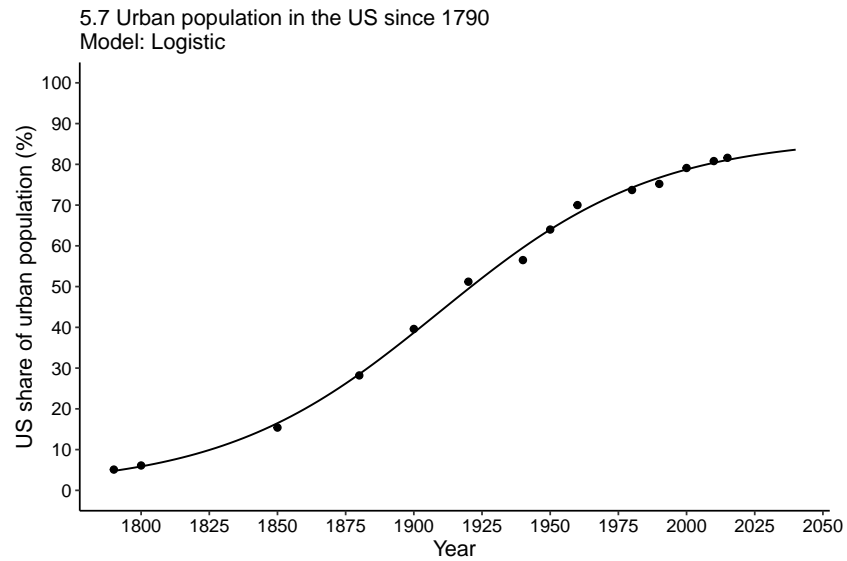
Dataset 5.7 (p. 338). Share of the urban population in the US since 1790. Plotted from data in USBC (1975) and subsequent volumes of US Statistical Abstract. With the logistic curve inflected already in 1910, its asymptotic value is 87.2%.

5.9.1 Tokyo's population: withinprefectural boundaries
Model: Logistic
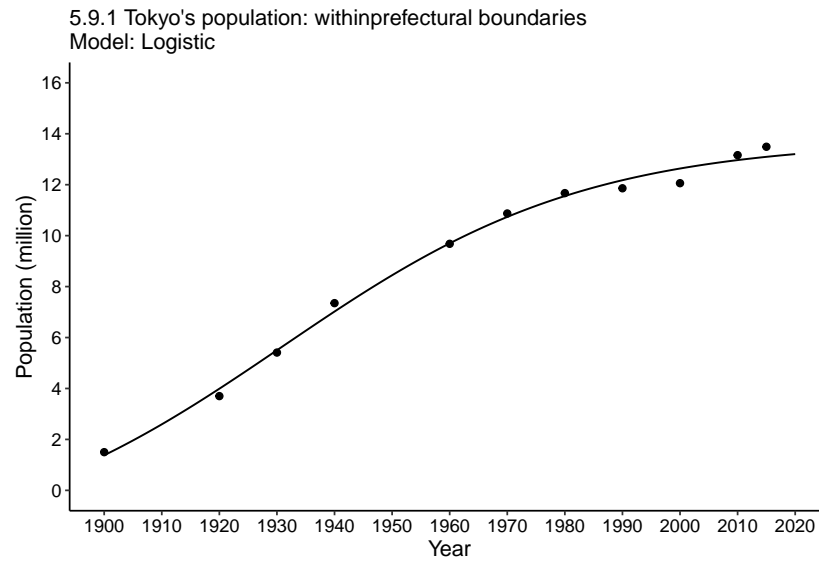
Dataset 5.9.1 (p. 349). Tokyo's population within prefectural boundaries (1900-2020) - a logistic curve with the inflection point in 1932 and asymptote at 13.8 million - and within the Tokyo Major Metropolitan Region since 1920: logistic curve infelction point in 1971 and asymptote at 38.8 million. Data from SB (1996) and TMG (2017).

5.9.2 Tokyo's population: within Major Metropolitan Region
Model: Logistic
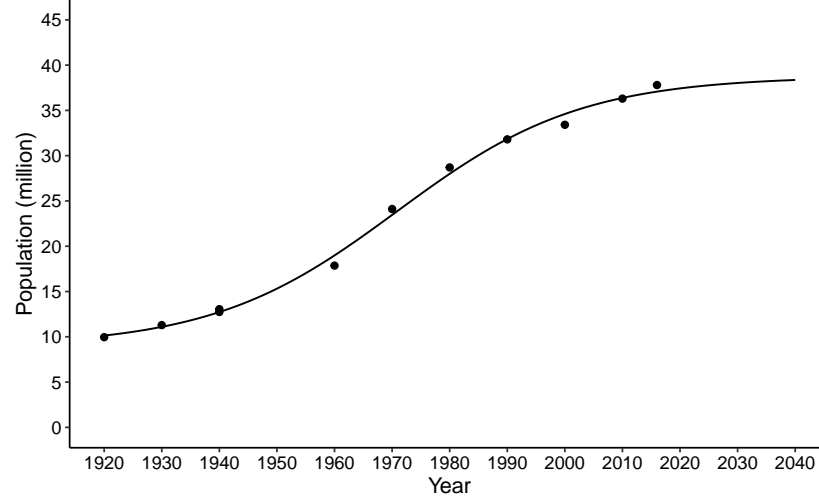
Dataset 5.9.2 (p. 349). Tokyo's population within prefectural boundaries (1900-2020) - a logistic curve with the inflection point in 1932 and asymptote at 13.8 million - and within the Tokyo Major Metropolitan Region since 1920: logistic curve infelction point in 1971 and asymptote at 38.8 million. Data from SB (1996) and TMG (2017).

5.11 Roman republic and empire, 509 BCE–117 CE and extrapolation
Model: Logistic
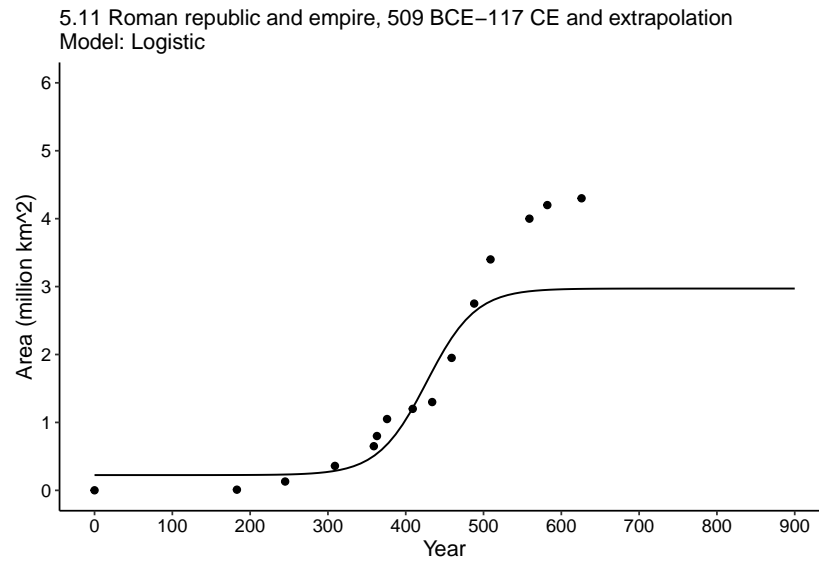
Dataset 5.11 (p. 372). Growth of the Roman republic and empire, 509 BCE-117 CE and the implied future trajectory based on a logistic fit. Data from Taagepera (1979) and author's area measurements based on Talbert (2000). Year zero is 509 BCE.

5.12 Roman republic and empire, 509 BCE–1453 CE
Model: Gaussian
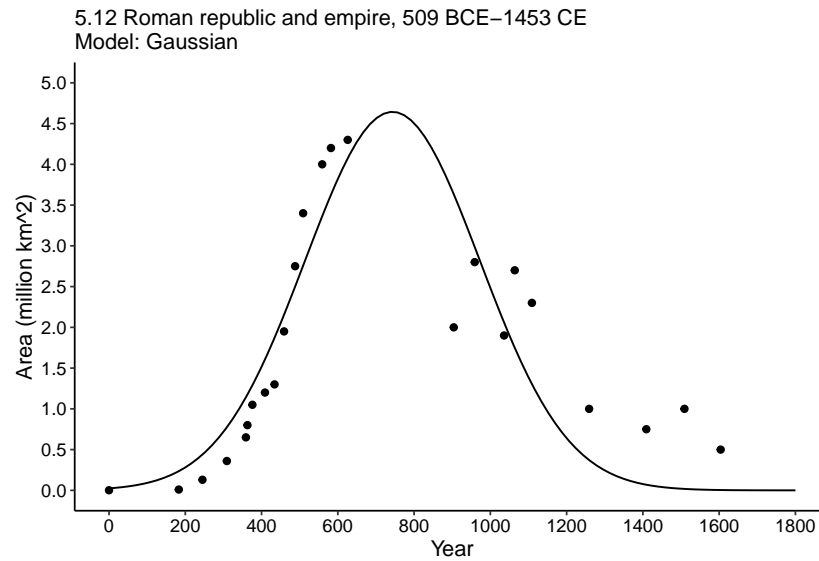
Dataset 5.12 (p. 373). Growth of the Roman republic and empire and the gradual retreat of Byzantium, 509 BCE-1453 CE. Data from Taagepera (1979) and author's area measurements based on Talbert (2000). Year zero is 509 BCE.

5.13.1 Global primary energy supply since 1800
Model: Logistic
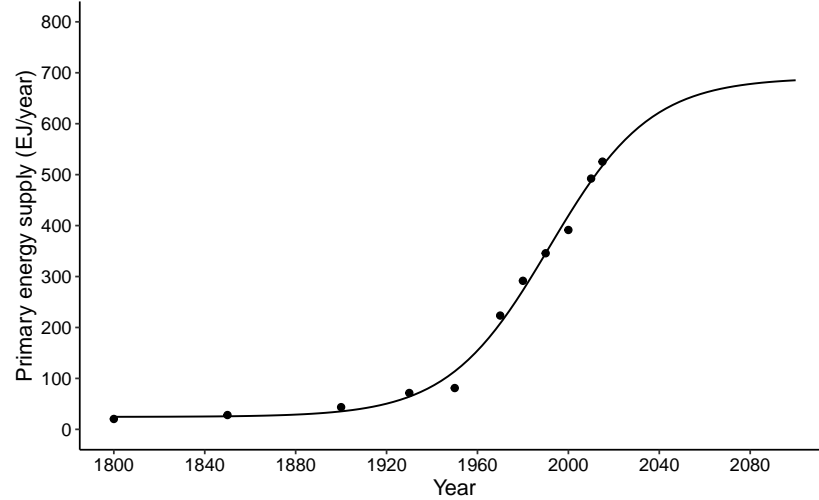
Dataset 5.13.1 (p. 379). Growth of global primary energy supply (including traditional biomass fuels) and fossil fuel and primary electricity supply since 1800. Data from Smil (2017b).

5.13.2 Fossil fuel and primary electricity supply
Model: Logistic
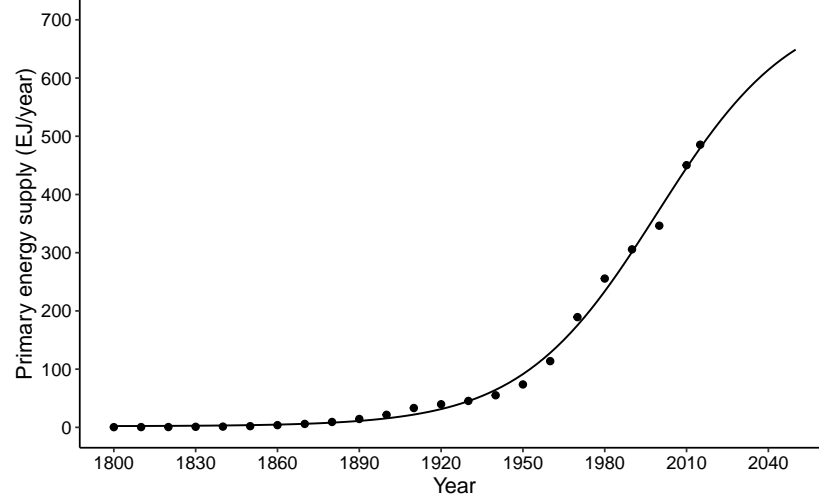
Dataset 5.13.2 (p. 379). Growth of global primary energy supply (including traditional biomass fuels) and fossil fuel and primary electricity supply since 1800. Data from Smil (2017b).
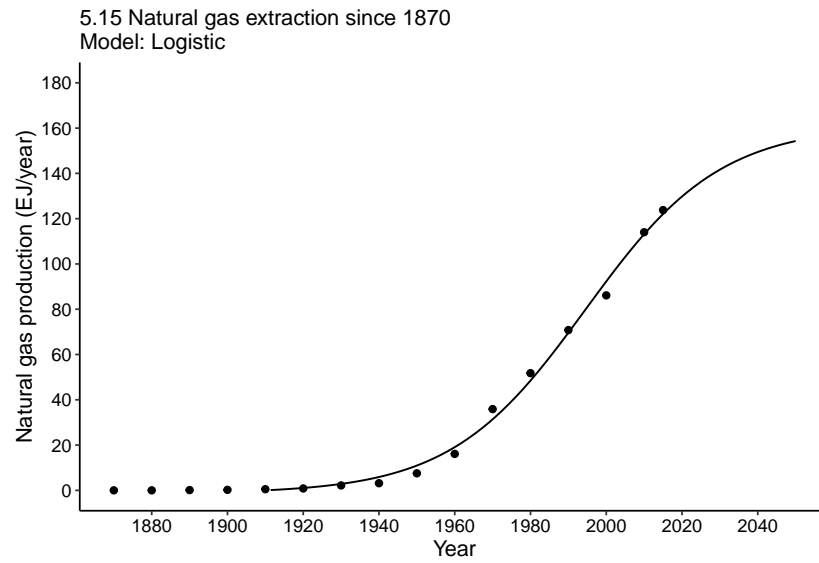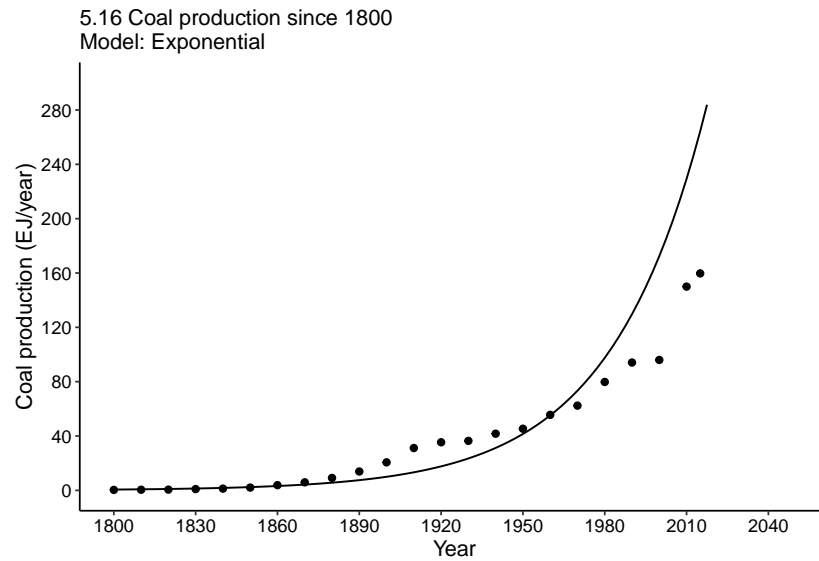
5.14 Crude oil production since 1870
Model: Logistic
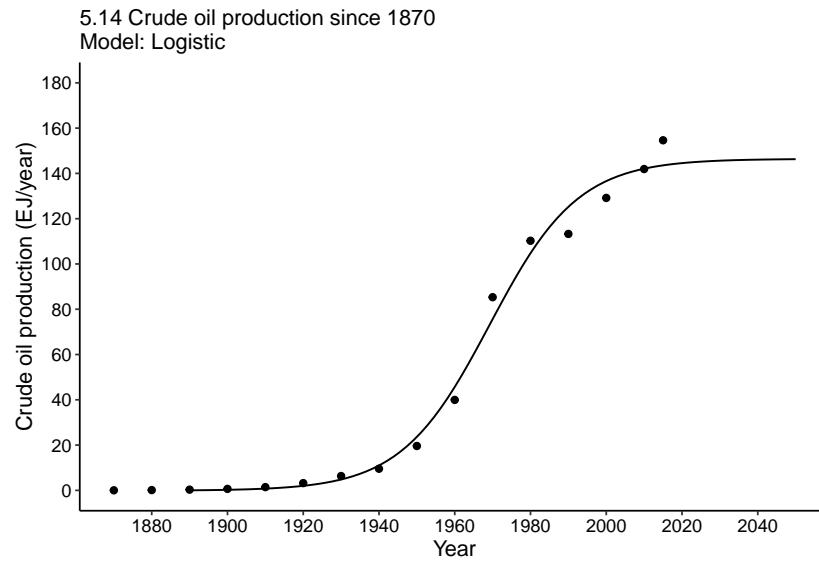
Dataset 5.14 (p. 380). Growth of global crude oil production since 1870. Data from Smil (2017b).

5.15 Natural gas extraction since 1870
Model: Logistic

**Dataset 5.15** (p. 381). Growth of global natural gas extraction since 1870. Data from Smil (2017b).

5.16 Coal production since 1800
Model: Exponential

Dataset 5.16 (p. 382). Growth of global coal produciton since 1800. While the two logistic curves for crude oil and natural gas provide a highly plausible indication of future development, coal's indicated trajectory will almost certainly diverge from the indicated trend. Data from Smil (2017b).

5.17 Electricity generation since 1900
Model: Logistic
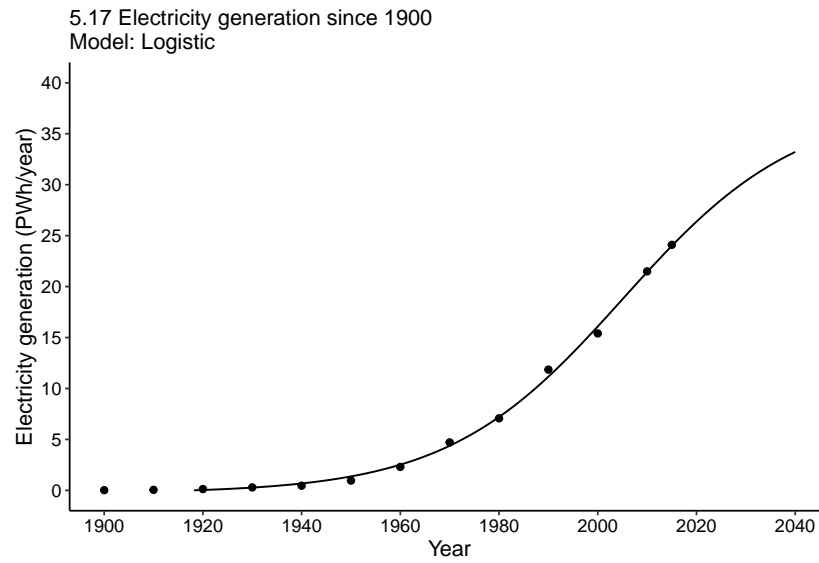
Dataset 5.17 (p. 383). Growth of global electricity generation since 1900. Data from Smil (2017b) and BP (2017).

5.18 Hydroelectricity generation since 1900
Model: Logistic
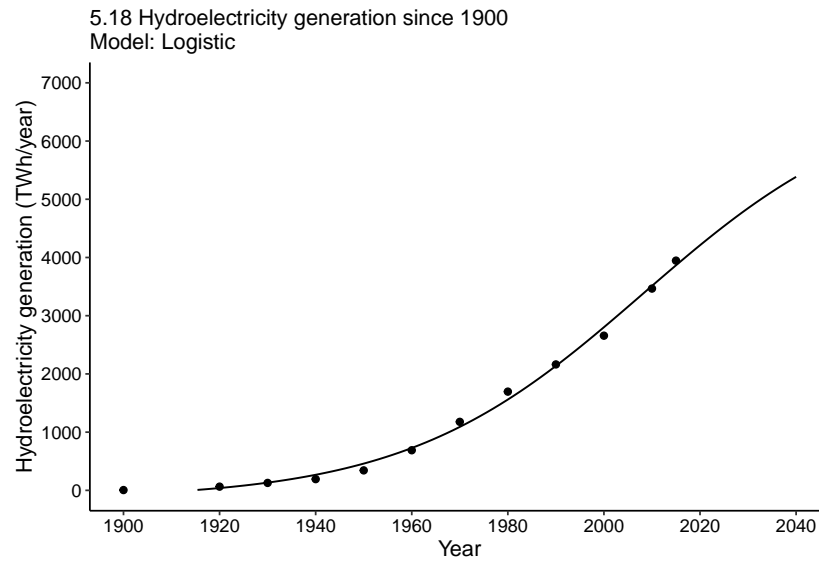
Dataset 5.18 (p. 384). Growth of global hydroelectricity generation since 1900. Data from Smil (2017b) and BP (2017).

5.19 Nuclear generation since 1960
Model: Logistic
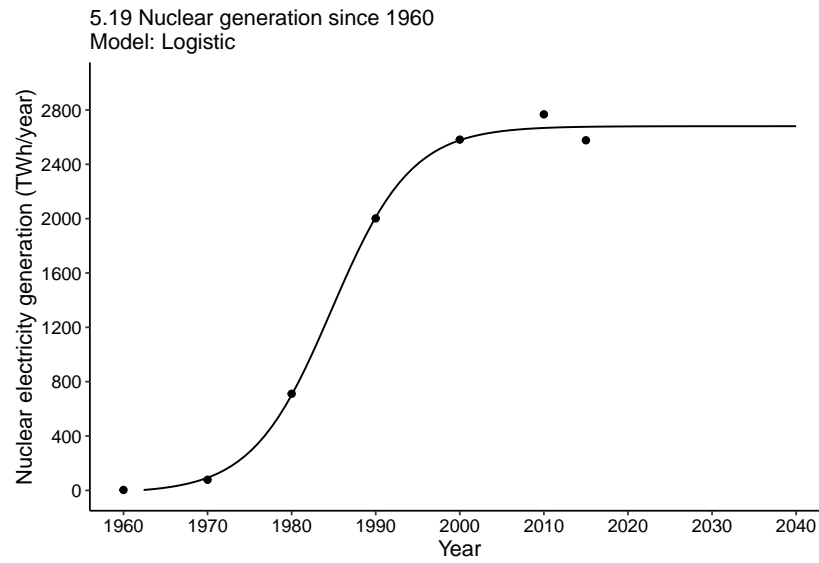
Dataset 5.19 (p. 385). Growth of global nuclear generation since 1960. Data from Smil (2017b) and BP (2017).

5.20 Global cropland 1700–2050
Model: Logistic
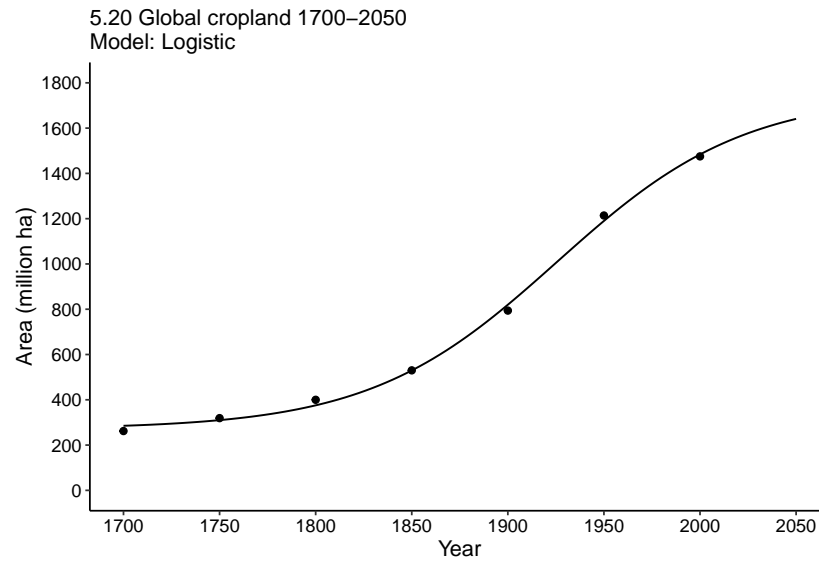
Dataset 5.20 (p. 387). Growth of global cropland, 1700-2050. Logistic curve had its inflection point in 1876 and the total is now less than 5% below the asymptote. Data from PBL (2010) and FAO (2018).

60

5.21 US cropland 1700–2050
Model: Logistic
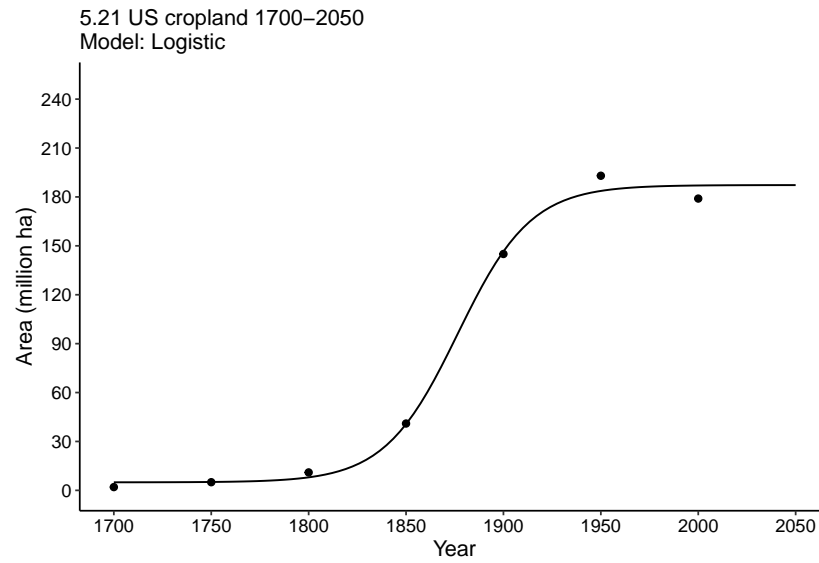
Dataset 5.21 (p. 387). Growth of US cropland, 1700-2050. Inflection point came already in 1876 and the recently cultivated area is very close to the asymptotic level. Data from PBL (2010) and FAO (2018).

5.22 Global grasslands
Model: Logistic
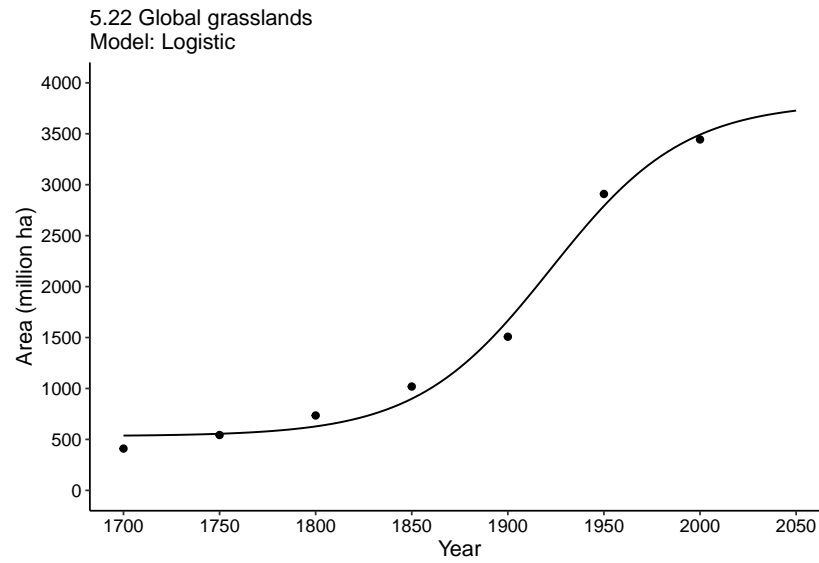
Dataset 5.22 (p. 388). Growth of global grasslands. The inflection year was in 1923 and the total area is now less than 10% from the asymptotic level. Data from PBL (2010) and FAO (2018).

5.23 Fertilizer production since 1913
Model: Logistic
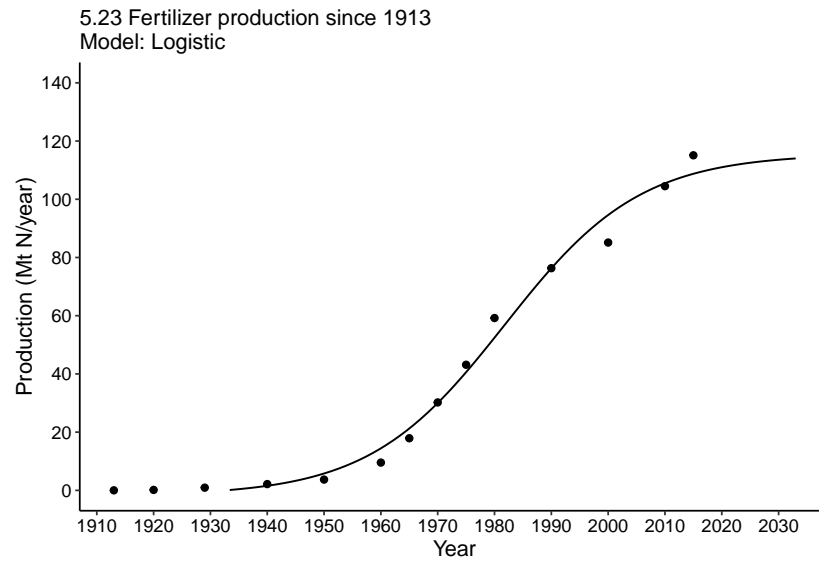
Dataset 5.23 (p. 390). Worldwide production of nitrogenous fertilizers since 1913. Another logistic curve that is very close to its asymptote. Data from Smil (2001) and FAO (2018).

5.24 Staple grain crop harvest since 1900
Model: Logistic
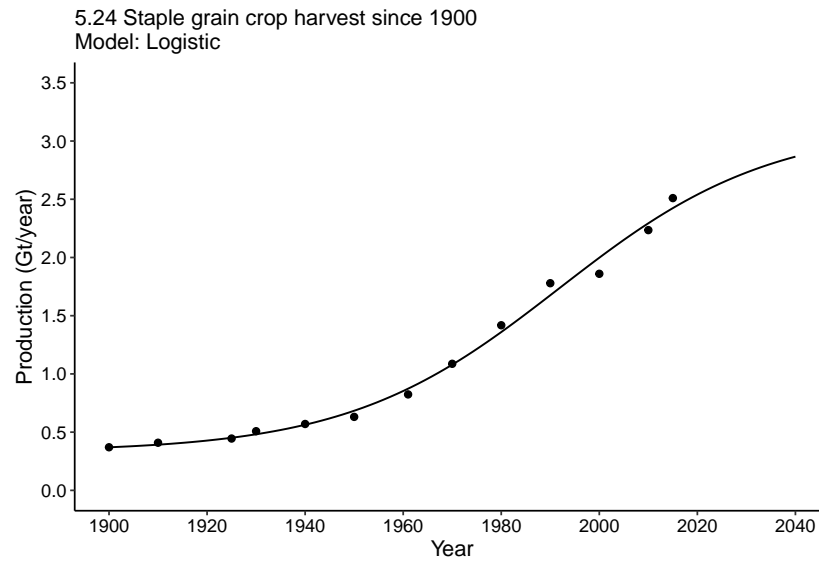
Dataset 5.24 (p. 391). Global harvest of staple grain crops since 1900. Data from Smil (2013a) and FAO (2018).

5.28 Semiconductor sales 1976–2016
Model: Logistic
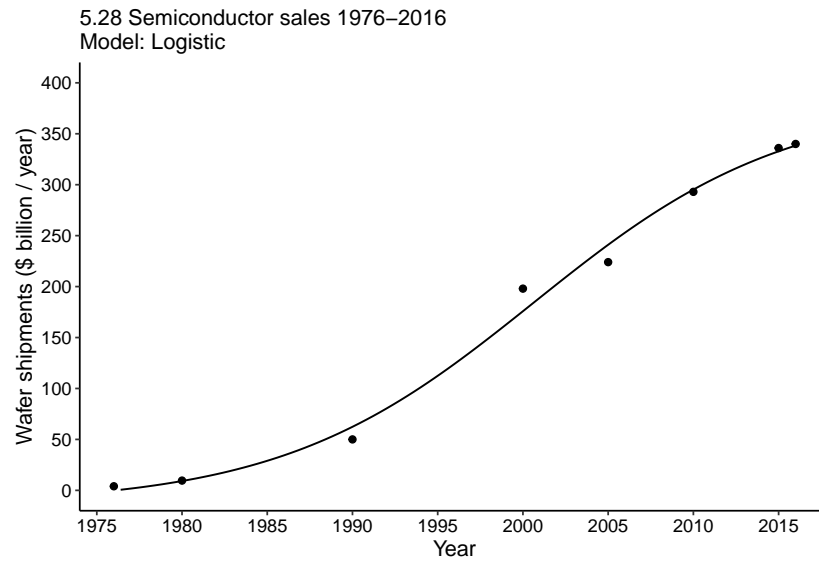
Dataset 5.28 (p. 398). Logistic curve of worldwide semiconductor sales, 1976-2016. Plotted from data in SIA (2017).

5.29.1 GDP (France)
Model: Logistic
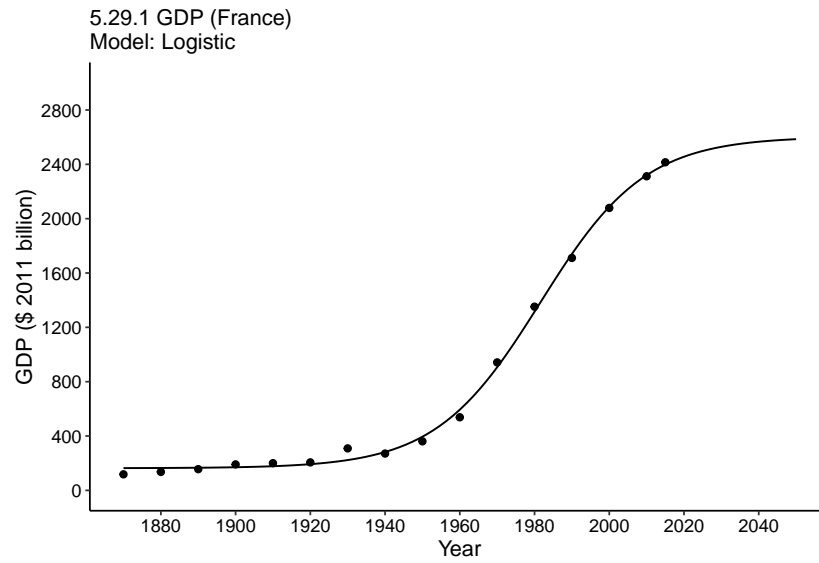
Dataset 5.29.1 (p. 412). Growth and logistic outlook of GDP (expression in international $2011) in four major economies - France, Japan, and the US since 1870, and China since 1950. Data from World Bank (2018).
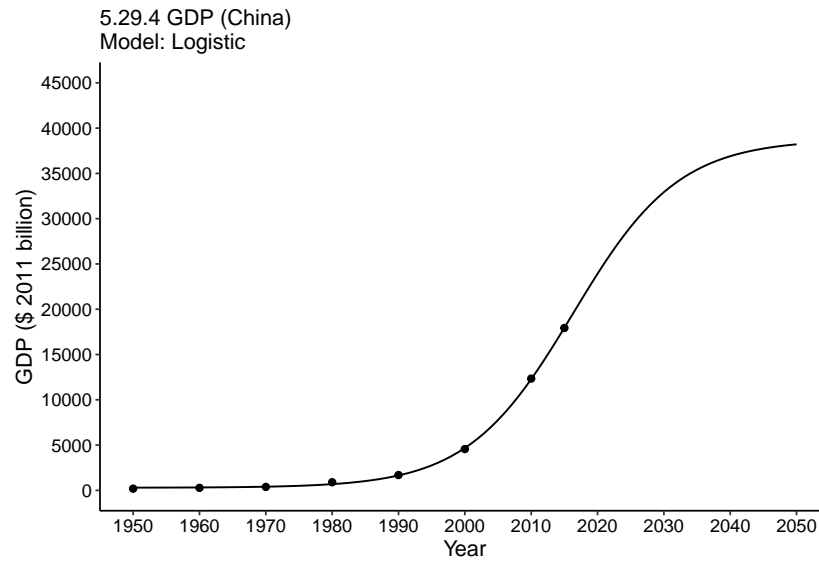
5.29.2 GDP (Japan)
Model: Logistic
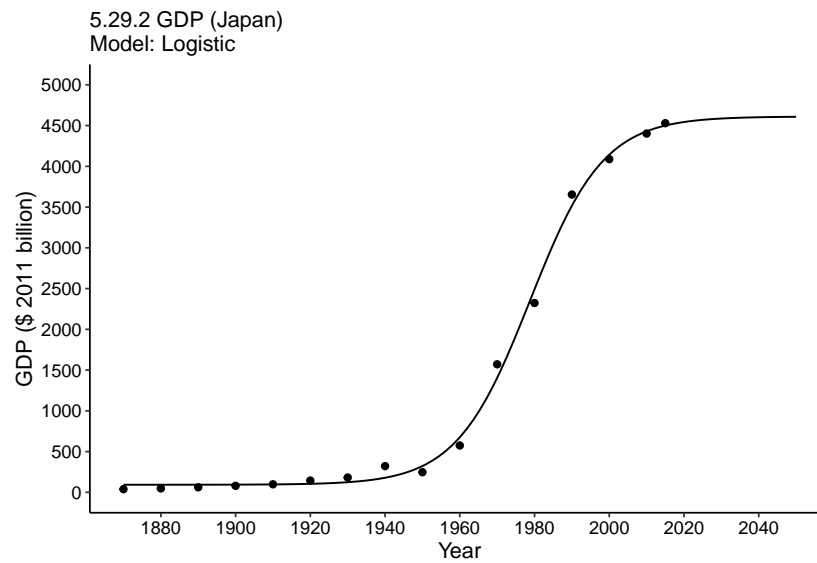
Dataset 5.29.2 (p. 412). Growth and logistic outlook of GDP (expression in international $2011) in four major economies - France, Japan, and the US since 1870, and China since 1950. Data from World Bank (2018).

5.29.3 GDP (US)
Model: Logistic
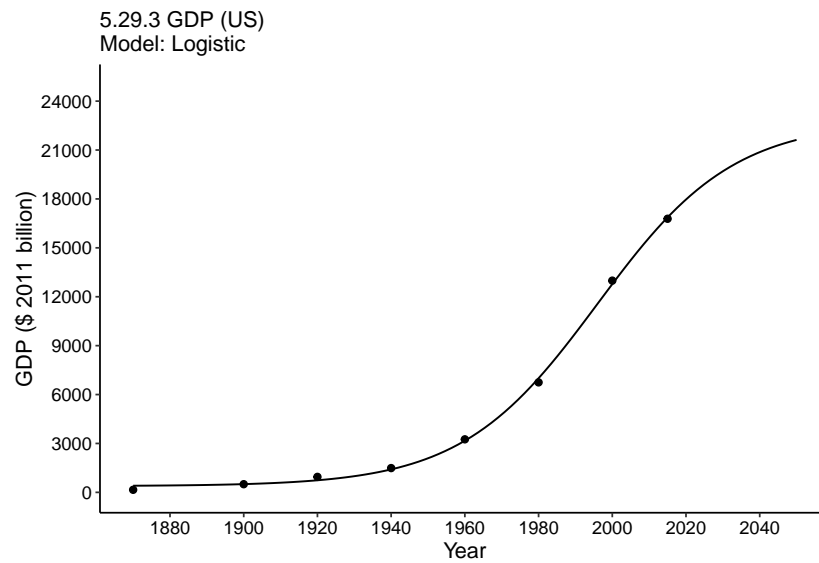
Dataset 5.29.3 (p. 412). Growth and logistic outlook of GDP (expression in international $2011) in four major economies - France, Japan, and the US since 1870, and China since 1950. Data from World Bank (2018).
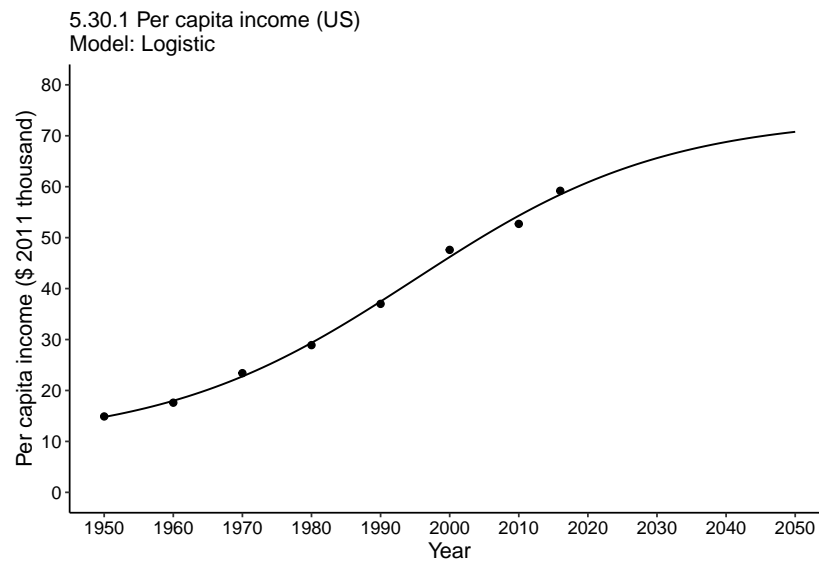
5.29.4 GDP (China)
Model: Logistic

Dataset 5.29.4 (p. 412). Growth and logistic outlook of GDP (expression in international $2011) in four major economies - France, Japan, and the US since 1870, and China since 1950. Data from World Bank (2018).
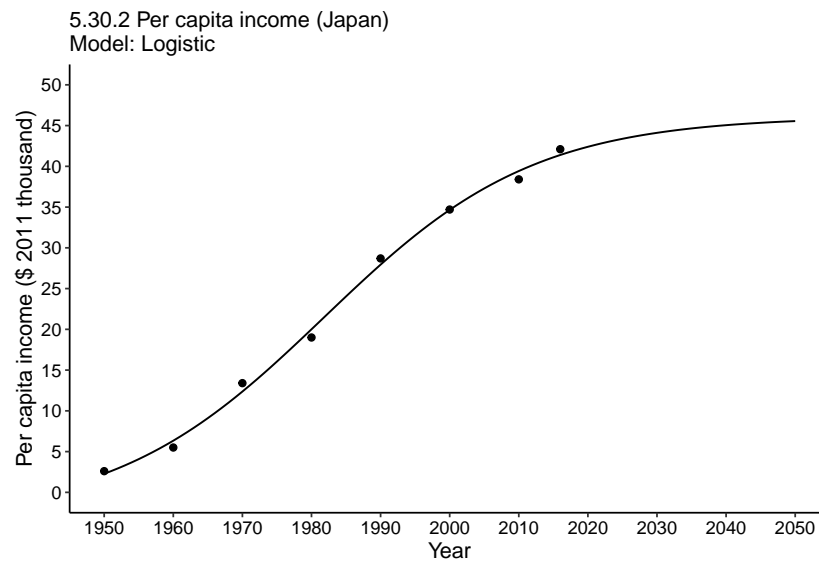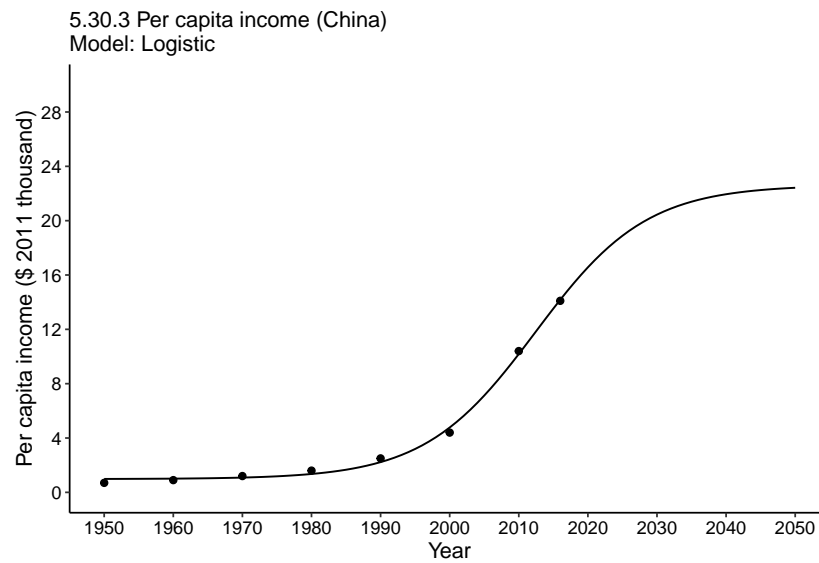
Dataset 5.30.1 (p. 418). Growth of per capita income in the US, Japan, and China since 1950. Data from World Bank (2018).

5.30.2 Per capita income (Japan)
Model: Logistic

Dataset 5.30.2 (p. 418). Growth of per capita income in the US, Japan, and China since 1950. Data from World Bank (2018).

5.30.3 Per capita income (China)
Model: Logistic

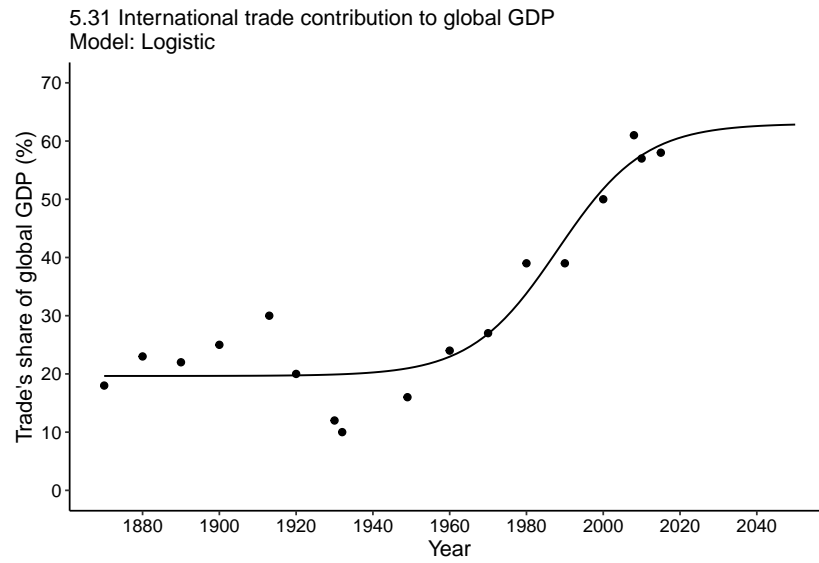**Dataset 5.30.3** (p. 418). Growth of per capita income in the US, Japan, and China since 1950. Data from World Bank (2018).

5.31 International trade contribution to global GDP
Model: Logistic

Dataset 5.31 (p. 431). Growing shares of international trade in the world economic product since 1870. Data from Klasing and Milionis (2014) and World Bank (2018).

6.1 US crude oil extraction forecast
Model: Gaussian
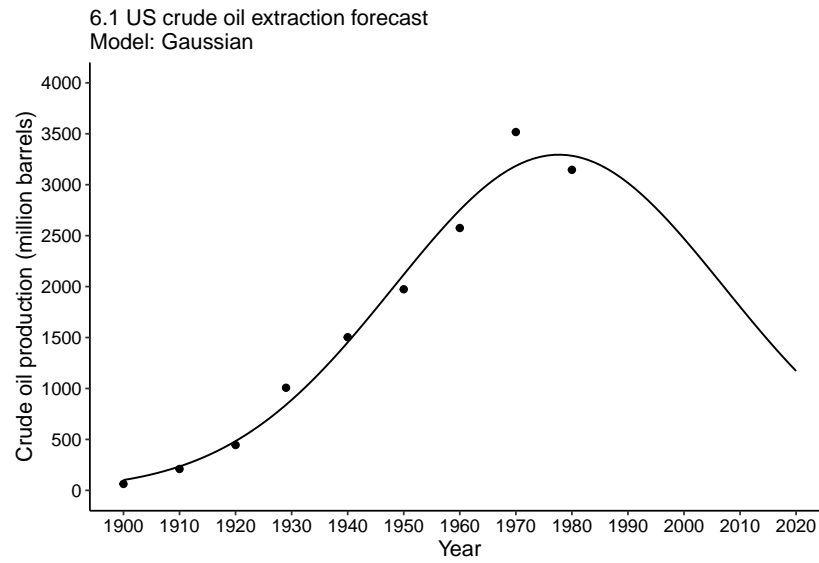
Dataset 6.1 (p. 452). US crude oil extraction forecast based on the 1900-1980 trajectory. Data from USBC (1975) and USEIA (2019).

6.2.1 Population of Greater London (a)
Model: Gaussian

Dataset 6.2.1 (p. 453). Population of Greater London: best fits based on 1801-1981 (a) and 1801-2001 (b) totals. Data from Morrey (1978) and GLA (2015).

75

6.2.2 Population of Greater London (b)
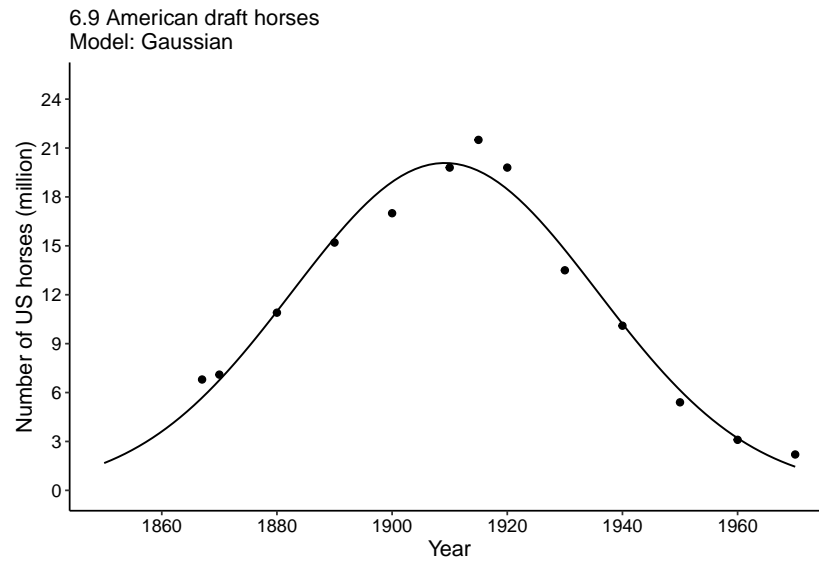Model: Logistic

Dataset 6.2.2 (p. 453). Population of Greater London: best fits based on 1801-1981 (a) and 1801-2001 (b) totals. Data from Morrey (1978) and GLA (2015).

6.7 British steel output
Model: Gaussian
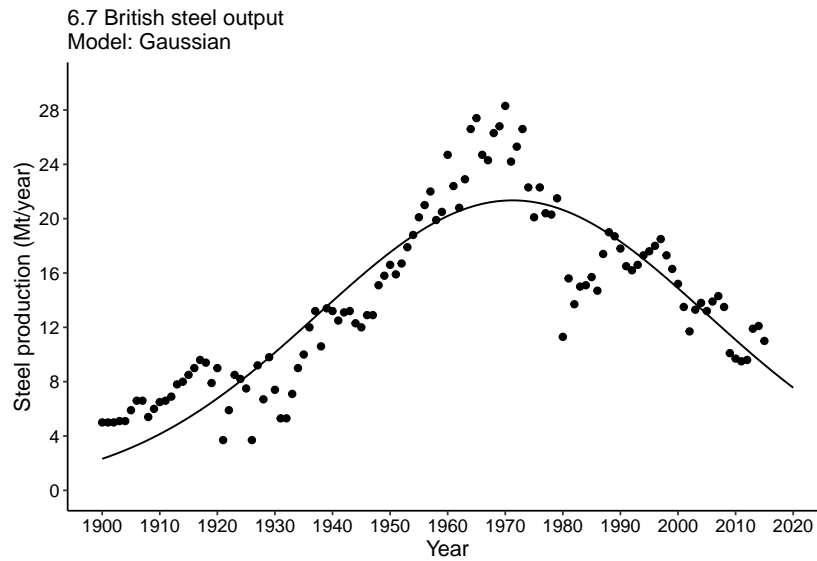
Dataset 6.7 (p. 464). British steel output, 1900-2020. Fluctuating output reflects economic downturns, expansions, and wars and hence the normal curve is not a particularly close fit (R^2 of 0.79). The production peak came in1970, with the 2015 output below the level attained first in 1936. Data from https://visual.ons.gov.uk/the-british-steel-industry-since-the-1970s

6.9 American draft horses
Model: Gaussian

Dataset 6.9 (p. 466). Number of American draft horses between 1850 and 1970 conforms closely to the normal curve trajectory. Data from USBC (1975).

6.10 US steam locomotives
Model: Gaussian
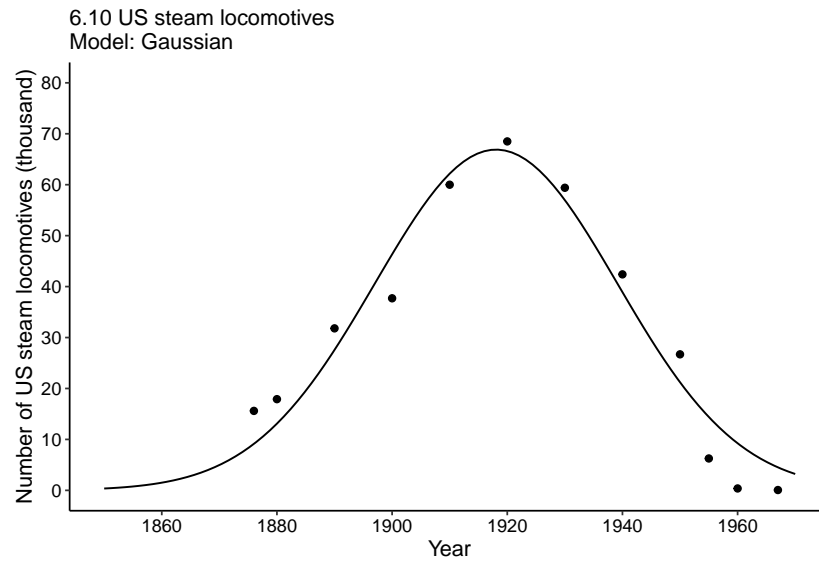
Dataset 6.10 (p. 467). Number of US steam locomotives: a very good Gaussian fit for the nine decades between 1876 and 1967. Data from USBC (1975).

6.11 US passenger car fleet
Model: Gaussian
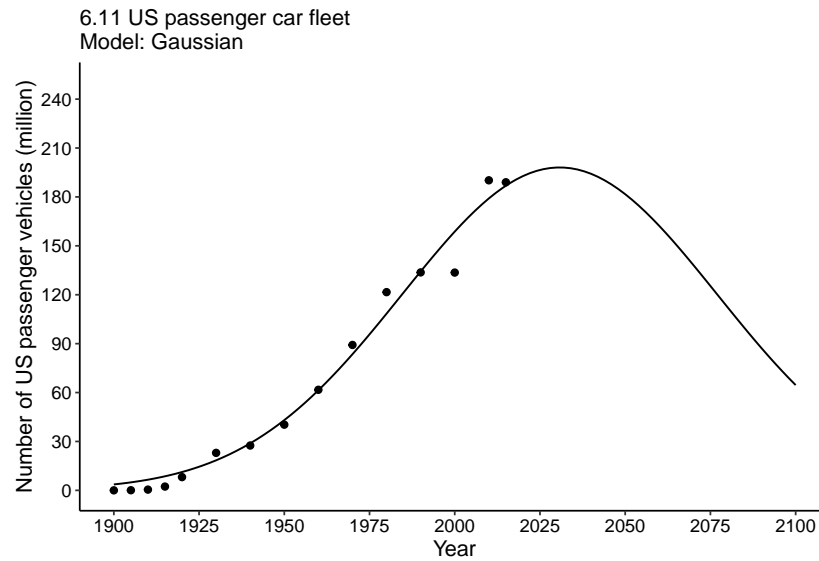
Dataset 6.11 (p. 467). The historical growth of the US passenger car fleet can be fitted quite well into a normal curve peaking around 2030. Data from USBC (1975) and from subsequent volumes of the US Statistical Abstract.