week1.R

ljh

2021-02-25

First we load the required packages:

```
library(tidyverse)
```

A general note on Rstudio projects

We are in an Rstudio project which means that:

"/data/dataset.csv"

. . . is equivalent to. . .

"C:/user/name/weird/idiosyncratic/directory/listings/project/data/dataset.csv"

A general note on nesting functions vs using pipes:

```
a <- head(cars)

b <- cars %>%
  head()

all.equal(a,b)
```

```
## [1] TRUE
```

```
a2 <- summary(head(cars))

b2 <- cars %>%
  head() %>%
  summary()

all.equal(a2,b2)
```

```
## [1] TRUE
```

Today's data is Tate artists/artworks

```
artwork <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_character(),
##   id = col_double(),
```

```
##   artistId = col_double(),
##   year = col_double(),
##   acquisitionYear = col_double(),
##   width = col_double(),
##   height = col_double(),
##   depth = col_double(),
##   thumbnailCopyright = col_logical()
## )
## i Use `spec()` for the full column specifications.
```

```
artists <- readr::read_csv("https://github.com/tategallery/collection/raw/master/artist_data.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   id = col_double(),
##   name = col_character(),
##   gender = col_character(),
##   dates = col_character(),
##   yearOfBirth = col_double(),
##   yearOfDeath = col_double(),
##   placeOfBirth = col_character(),
##   placeOfDeath = col_character(),
##   url = col_character()
## )
```

Task: put together a general report including some visualisations; include something that requires merging the two files

```
# Cross-tabulation
table(artists$gender, artists$yearOfBirth, useNA = "ifany")
```

```
##
##          1497 1500 1530 1531 1540 1547 1551 1560 1561 1572 1577 1580 1582 1585 1590 1593 1594 1599 16
##   Female    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
##   Male      1    0    0    1    2    1    1    1    1    1    2    1    1    1    1    1    1    1
##
##          1609 1611 1615 1617 1618 1621 1626 1627 1628 1629 1630 1631 1632 1633 1635 1639 1640 1641 16
##   Female    0    0    0    0    0    0    0    0    0    0    0    0    0    1    0    0    0    0
##   Male      0    2    0    1    1    1    1    1    1    1    2    1    1    3    1    1    3    1
##
##          1646 1647 1652 1656 1659 1660 1662 1667 1674 1675 1679 1680 1681 1682 1684 1685 1686 1689 16
##   Female    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
##   Male      2    1    1    1    1    4    1    1    0    1    1    1    1    2    1    3    1    1
##
##          1696 1697 1698 1699 1700 1701 1702 1703 1705 1707 1708 1710 1711 1713 1714 1715 1716 1717 17
##   Female    0    0    0    0    1    0    0    0    0    0    0    0    0    0    0    0    0    0
##   Male      1    3    1    2    6    1    5    1    1    1    2    4    1    3    2    1    1    1
##
##          1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 17
##   Female    0    0    1    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
##   Male      3    2    2    2    4    3    3    6    1    2    4    6    4    1    1    4    2    6
##
##          1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 17
##   Female    0    0    0    0    1    0    2    0    0    0    0    0    0    0    0    0    0    0
##   Male      7    2    4    3    3    3    6    4    4    4    4    6    5    4    3    5    5    2
```

```
##
##          1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 17
##   Female    1    0    0    0    0    0    0    1    0    0    1    1    1    1    0    0    0    1
##   Male      3    2   10    3    9    5    1    4    6    3    8    2    4    6    2    9    7   12
##
##          1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 18
##   Female    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
##   Male      7    8    6    7    6    4    7    5    7    4    4    3    3    6   10    3    7   10
##
##          1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824 18
##   Female    0    0    1    1    0    0    0    0    0    0    0    1    0    0    0    0    0    0
##   Male      6    3    1    7    5    7    7    3    5    3   10    4    8    5    6    7    4    8
##
##          1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844 1845 18
##   Female    1    2    0    1    1    1    0    0    0    0    0    0    1    1    0    0    3    0
##   Male     11    6    7    6    7    7    6    8    8    4    5    6   10    9    6    4    3    9
##
##          1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 18
##   Female    0    0    1    0    0    0    2    1    0    1    1    0    2    0    1    1    1    3
##   Male     13    9    7    5    4    8    8   10   11    6   11   14   16   14   12   12   11   10
##
##          1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 18
##   Female    2    1    1    1    2    2    2    1    5    3    1    3    1    1    0    4    1    1
##   Male     12    9   15   10   14   10   13    8   16    9   13    9    8   17   11   12   20   15
##
##          1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 19
##   Female    1    4    3    3    4    1    1    2    4    4    4    4    5    7    2    7    4    4
##   Male     12   12   13   20   13   11   12   16   14   14   16   13   19   24   19   15   17   14
##
##          1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 19
##   Female    6    1    7    3    3    3    1    3    1    4    3    6    4    2    3    4    3    3
##   Male     18   21   14   18   10   10   15   14   20   24   24   30   22   33   21   26   41   19
##
##          1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 19
##   Female    3    4    4    6    4    8    4    6   10    4    2    8    5    4    5    4    5    4
##   Male     24   23   25   42   24   33   29   23   30   17   20   20   26   21   23   21   21   12
##
##          1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 19
##   Female    5    4    9    4    5    7    5    7    5   14    6   11   10   15    7    7    5    7
##   Male     15   15   13   12   16   15   15   21   19   13   18   16   21   29   16   15   15   15
##
##          1975 1976 1977 1978 1979 1980 1981 1982 1985 1986 1988 1992 1996 1999 2004 <NA>
##   Female    5    1    8    2    2    2    1    0    0    0    0    0    0    0    0    9
##   Male     10    3   10    4    4    0    0    3    0    0    1    0    0    0    1   23
##  [ reached getOption("max.print") -- omitted 1 row ]
```

```
#Simplify the creditLine and medium variables

artwork_simpl <- artwork %>%
  mutate(credit_simple = factor(case_when(str_detect(creditLine, "Presented") ~ "Presented",
                                          str_detect(creditLine, "Purchased") ~ "Purchased",
                                          str_detect(creditLine, "Bequeathed") ~ "Bequeathed"))) %>%
  separate(medium,into= paste0("word",seq(1,10,1)), sep = " ") %>%
  drop_na(credit_simple)
```

```
## Warning: Expected 10 pieces. Additional pieces discarded in 427 rows [1, 1062, 1065, 1067, 1527, 176
## 1806, 1906, 1911, 1912, 1913, 1914, 2243, 2384, 2416, 2440, 2866, 2889, ...].
```

```
## Warning: Expected 10 pieces. Missing pieces filled with `NA` in 62204 rows [2, 3, 4, 5, 6, 7, 8, 9,
## 14, 15, 16, 17, 18, 19, 20, 21, ...].
```

```
# Visualise the association between year and acquisition year
ggplot(data = artwork_simpl, aes(x=year, y=acquisitionYear, colour=credit_simple, size=width/100)) +
  geom_point()
```

```
## Warning: Removed 6047 rows containing missing values (geom_point).
```