DplyrTidyrLab

Sarah Grant

Exercise 1

Question 1

```
The data set contains 1926 unique songs.
```

```
The data set contains 835 unique artists.
```

```
The data set contains 59 unique musical genres.
```

Question 2

Table 1: Number of Artists per Year

Year	Number of Distinct Artists
1998	1
1999	30
2000	58
2001	77
2002	56
2003	64
2004	65
2005	69
2006	63
2007	66
2008	63
2009	61
2010	62
2011	69
2012	67

Year	Number of Distinct Artists
2013	64
2014	70
2015	69
2016	75
2017	86
2018	81
2019	73
2020	3

Question 3

```
The most popular artist is Rihanna with 23 songs.
```

Question 4

Below is the table showing the minimum, maximum, mean and median tempo as well as the number of songs, for each musical genre:

Table 2: Tempo Statistics by Genre

Genre	MinTempo	MaxTempo	MeanTempo	MedianTempo	Number of songs
Dance/Electronic	75.255	179.642	125.5075	126.0410	40
Folk/Acoustic, pop	94.931	128.945	111.9380	111.9380	2
Folk/Acoustic, rock	84.192	84.192	84.1920	84.1920	1
Folk/Acoustic, rock, pop	138.585	138.585	138.5850	138.5850	1
R&B	71.815	170.661	106.9248	100.4600	13
World/Traditional, Folk/Acoustic	82.803	82.803	82.8030	82.8030	1
World/Traditional, hip hop	98.077	101.993	100.0350	100.0350	2
World/Traditional, pop	108.102	108.102	108.1020	108.1020	1
World/Traditional, pop, Folk/Acoustic	100.380	104.833	102.6065	102.6065	2
World/Traditional, rock	96.000	140.083	118.0415	118.0415	2
World/Traditional, rock, pop	132.013	139.048	135.5305	135.5305	2
country	103.055	205.570	138.1508	136.0020	9
country, latin	96.055	96.055	96.0550	96.0550	1
easy listening	157.920	157.920	157.9200	157.9200	1
hip hop	64.934	179.974	116.9894	111.6795	120

Genre	MinTempo	MaxTempo	MeanTempo	MedianTempo	Number of songs
hip hop, Dance/Electronic	95.948	190.151	135.4297	131.0500	15
hip hop, R&B	100.215	151.181	121.1220	111.9700	3
hip hop, country	97.984	97.984	97.9840	97.9840	1
hip hop, latin, Dance/Electronic	171.993	171.993	171.9930	171.9930	1
hip hop, pop	73.003	203.911	118.9619	119.9750	265
hip hop, pop, Dance/Electronic	72.022	196.093	120.8555	126.0620	75
hip hop, pop, R&B	60.019	203.862	115.2649	107.3310	232
hip hop, pop, R&B, Dance/Electronic	82.820	127.901	103.9113	101.0130	3
hip hop, pop, R&B, latin	82.331	100.010	91.1705	91.1705	2
hip hop, pop, country	129.370	129.370	129.3700	129.3700	1
hip hop, pop, latin	89.661	180.184	127.2119	127.0265	14
hip hop, pop, rock	84.858	179.999	123.1123	125.2500	9
hip hop, rock, pop	90.052	90.052	90.0520	90.0520	1
latin	90.013	198.075	121.6049	97.0620	15
metal	79.012	147.387	106.2089	101.9680	9
pop	65.043	195.685	120.7527	119.9535	405
pop, Dance/Electronic	84.878	198.065	123.4655	124.0595	213
pop, Folk/Acoustic	76.026	171.790	118.3595	109.9505	8
pop, R&B	68.942	210.851	117.4385	112.5110	169
pop, R&B, Dance/Electronic	84.021	176.051	112.0338	104.0865	6
pop, R&B, easy listening	108.984	108.984	108.9840	108.9840	1
pop, country	97.865	147.905	130.5087	136.9250	8
pop, easy listening, Dance/Electronic	135.099	135.099	135.0990	135.0990	1
pop, easy listening, jazz	82.168	127.831	104.9995	104.9995	2
pop, latin	79.997	177.833	113.5903	104.2540	28
pop, rock	77.967	176.667	121.0976	119.0095	26
pop, rock, Dance/Electronic	87.016	189.857	133.9808	135.9875	12
pop, rock, Folk/Acoustic	102.961	112.960	107.9605	107.9605	2
pop, rock, metal	82.952	155.827	128.9358	134.7165	14
rock	74.989	199.935	129.5312	123.6960	57
rock, Dance/Electronic	127.988	127.988	127.9880	127.9880	1
rock, Folk/Acoustic, easy listening	122.979	122.979	122.9790	122.9790	1
rock, Folk/Acoustic, pop	80.529	80.529	80.5290	80.5290	1
rock, R&B, Folk/Acoustic, pop	105.987	105.987	105.9870	105.9870	1
rock, blues	123.904	141.933	132.9185	132.9185	2
rock, blues, latin	97.911	127.981	112.9460	112.9460	2
rock, classical	81.663	81.663	81.6630	81.6630	1

Genre	MinTempo	MaxTempo	MeanTempo	MedianTempo	Number of songs
rock, easy listening	114.999	114.999	114.9990	114.9990	1
rock, metal	89.342	187.961	127.3922	120.0555	36
rock, pop	68.976	184.086	123.8996	124.9700	39
rock, pop, Dance/Electronic	113.049	181.994	135.7678	127.4480	8
rock, pop, metal	126.115	152.034	140.2785	141.4825	4
rock, pop, metal, Dance/Electronic	105.013	105.013	105.0130	105.0130	1
set()	68.507	184.819	120.1329	126.9620	22

Question 5

Below is the simple dataframe displaying the mean liveness and mean danceability per year:

```
# A tibble: 23 x 3
    year mean_liveness mean_danceability
   <int>         <dbl>             <dbl>
 1  1998         0.18              0.727
 2  1999         0.166             0.669
 3  2000         0.181             0.690
 4  2001         0.174             0.674
 5  2002         0.193             0.675
 6  2003         0.163             0.665
 7  2004         0.180             0.697
 8  2005         0.188             0.673
 9  2006         0.198             0.661
10  2007         0.184             0.631
# i 13 more rows
```

Just for a more "aesthetically pleasing" depiction, below is the dataframe knitted into a table:

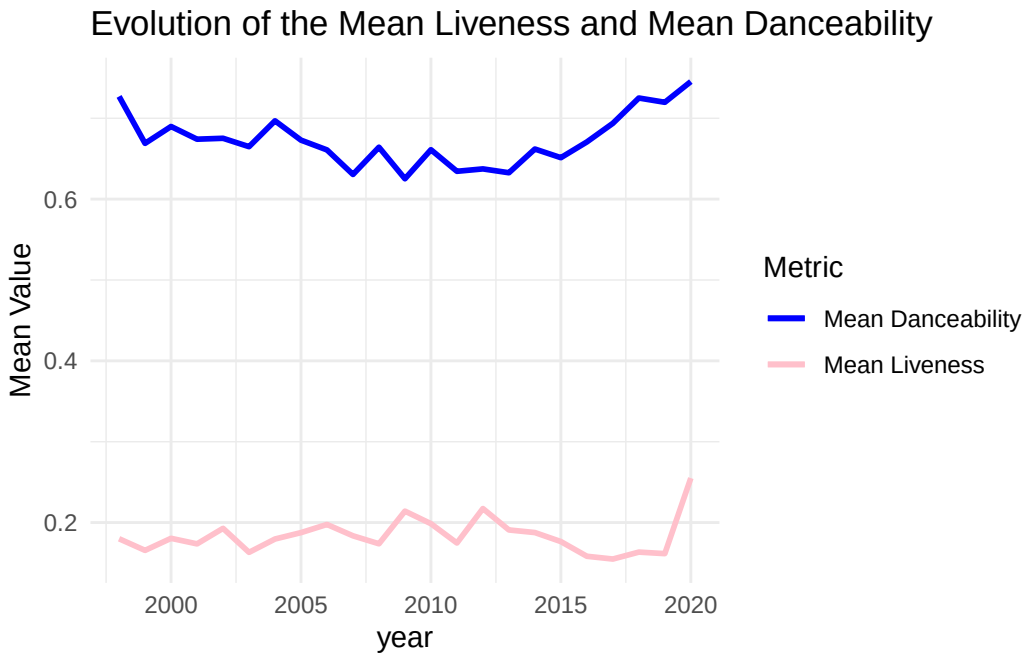
Table 3: Mean Liveness and Mean Danceability per Year

Year	Mean Liveness	Mean Danceability
1998	0.1800000	0.7270000
1999	0.1656000	0.6689737
2000	0.1805216	0.6898243
2001	0.1736685	0.6741296
2002	0.1928467	0.6752444

Year	Mean Liveness	Mean Danceability
2003	0.1631701	0.6648763
2004	0.1796552	0.6968333
2005	0.1875663	0.6729231
2006	0.1976642	0.6608632
2007	0.1836223	0.6305213
2008	0.1737474	0.6641856
2009	0.2141405	0.6251667
2010	0.1986150	0.6610748
2011	0.1747636	0.6344040
2012	0.2173391	0.6373652
2013	0.1908663	0.6326404
2014	0.1876317	0.6619327
2015	0.1765152	0.6512626
2016	0.1583172	0.6707071
2017	0.1549054	0.6937387
2018	0.1634561	0.7250374
2019	0.1615809	0.7197640
2020	0.2550000	0.7453333

Question 6

Below is graphical depiction of the temporal evoloution of both the mean annual liveness and the mean annual danceability.



Side note for professor: This graph is not too "80s vibes" in terms of colour choice I hope!

Exercise 2

Question 1

Below we can see the median admission grade for each combination of Target variable and Marital Status:

Target	Marital status	Median admission grade
Dropout	single	123.35
Dropout	married	126.50
Dropout	divorced	126.50
Dropout	widower	129.40
Dropout	facto union	119.40
Dropout	legally separated	112.50
Graduate	single	127.30
Graduate	married	130.00
Graduate	divorced	126.00
Graduate	widower	170.00
Graduate	facto union	120.00
Graduate	legally separated	114.80

Target	Marital status	Median admission grade
Enrolled	single	124.05
Enrolled	married	122.95
Enrolled	divorced	130.20
Enrolled	widower	151.75
Enrolled	facto union	119.70
Enrolled	legally separated	119.00

Question 2

The dataframe in Question 1 isn't the best way that we can show this... Below is the transformation of the previous dataframe with each row corresponding to a specific marital status (which is stated), while the other columns contain the corresponding median grade:

Marital status	Dropout	Graduate	Enrolled
single	123.35	127.3	124.05
married	126.50	130.0	122.95
divorced	126.50	126.0	130.20
widower	129.40	170.0	151.75
facto union	119.40	120.0	119.70
legally separated	112.50	114.8	119.00

Question 3

Below is the dataframe showing the conditional median of all variables related to "Curricular units" grouped by gender:

Gender	Curricular units 1st sem (credited)	Curricular units 1st sem (enrolled)	Curricular units 1st sem (evaluations)	Curricular units 1st sem (approved)	Curricular units 1st sem (grade)	Curricular units 1st sem (without evaluations)	Curricular units 2nd sem (credited)	Curricular units 2nd sem (enrolled)	Curricular units 2nd sem (evaluations)	Curricular units 2nd sem (approved)	Curricular units 2nd sem (grade)	Curricular units 2nd sem (without evaluations)
Male	0	6	8	4	11.83333	0	0	6	8	4	11.63604	0
Female	0	6	8	6	12.50000	0	0	6	8	5	12.50000	0

Question 4

Below we can see the transformed data which is more readable, displayed in a knitted table:

Table 7: Conditional Median of Curricular Units by Gender

Units	Male	Female
Curricular units 1st sem (approved)	4.00000	6.0
Curricular units 1st sem (credited)	0.00000	0.0
Curricular units 1st sem (enrolled)	6.00000	6.0
Curricular units 1st sem (evaluations)	8.00000	8.0
Curricular units 1st sem (grade)	11.83333	12.5
Curricular units 1st sem (without evaluations)	0.00000	0.0
Curricular units 2nd sem (approved)	4.00000	5.0
Curricular units 2nd sem (credited)	0.00000	0.0
Curricular units 2nd sem (enrolled)	6.00000	6.0
Curricular units 2nd sem (evaluations)	8.00000	8.0
Curricular units 2nd sem (grade)	11.63604	12.5
Curricular units 2nd sem (without evaluations)	0.00000	0.0