**数据分析报告**

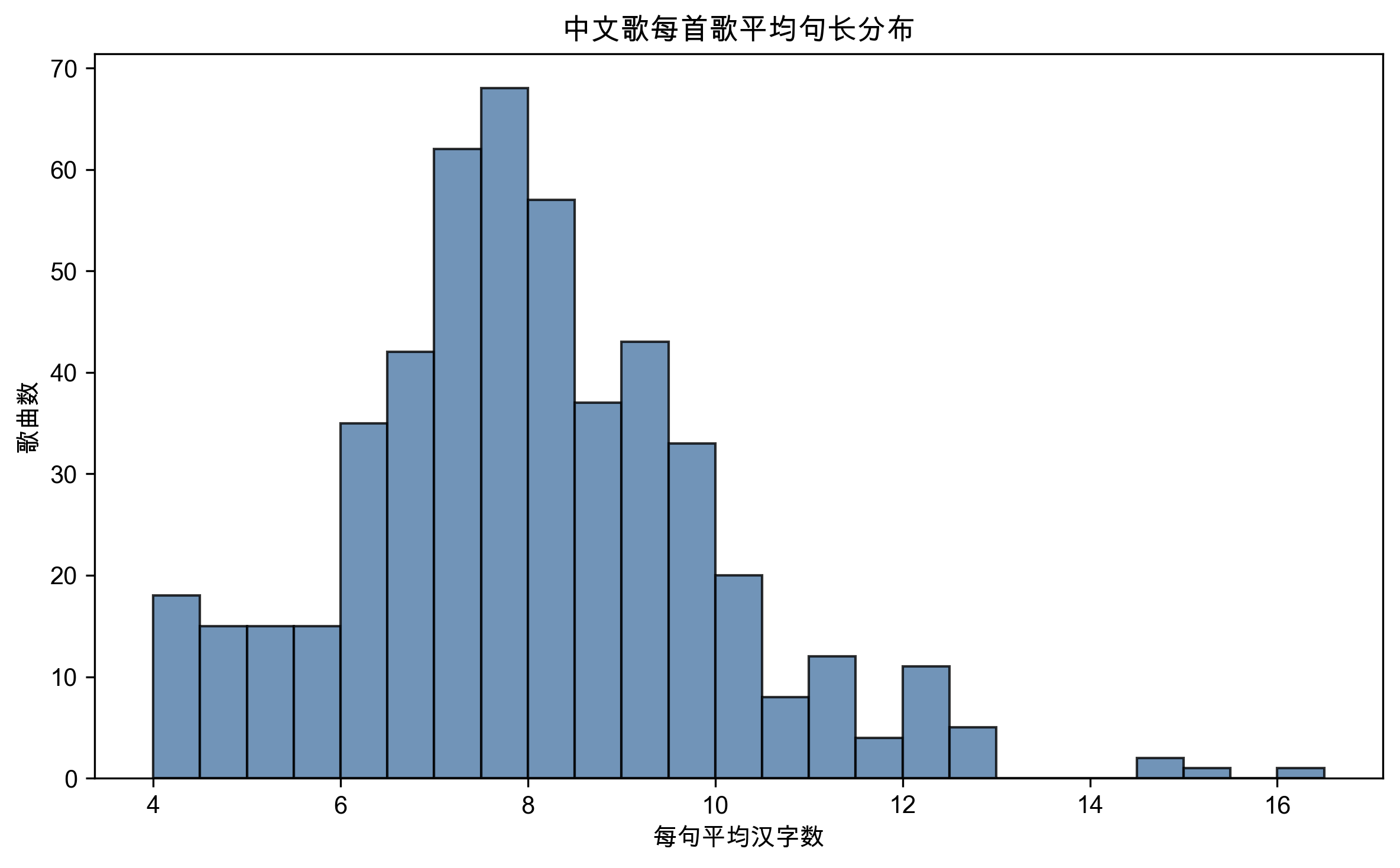
2024010766 陈家康

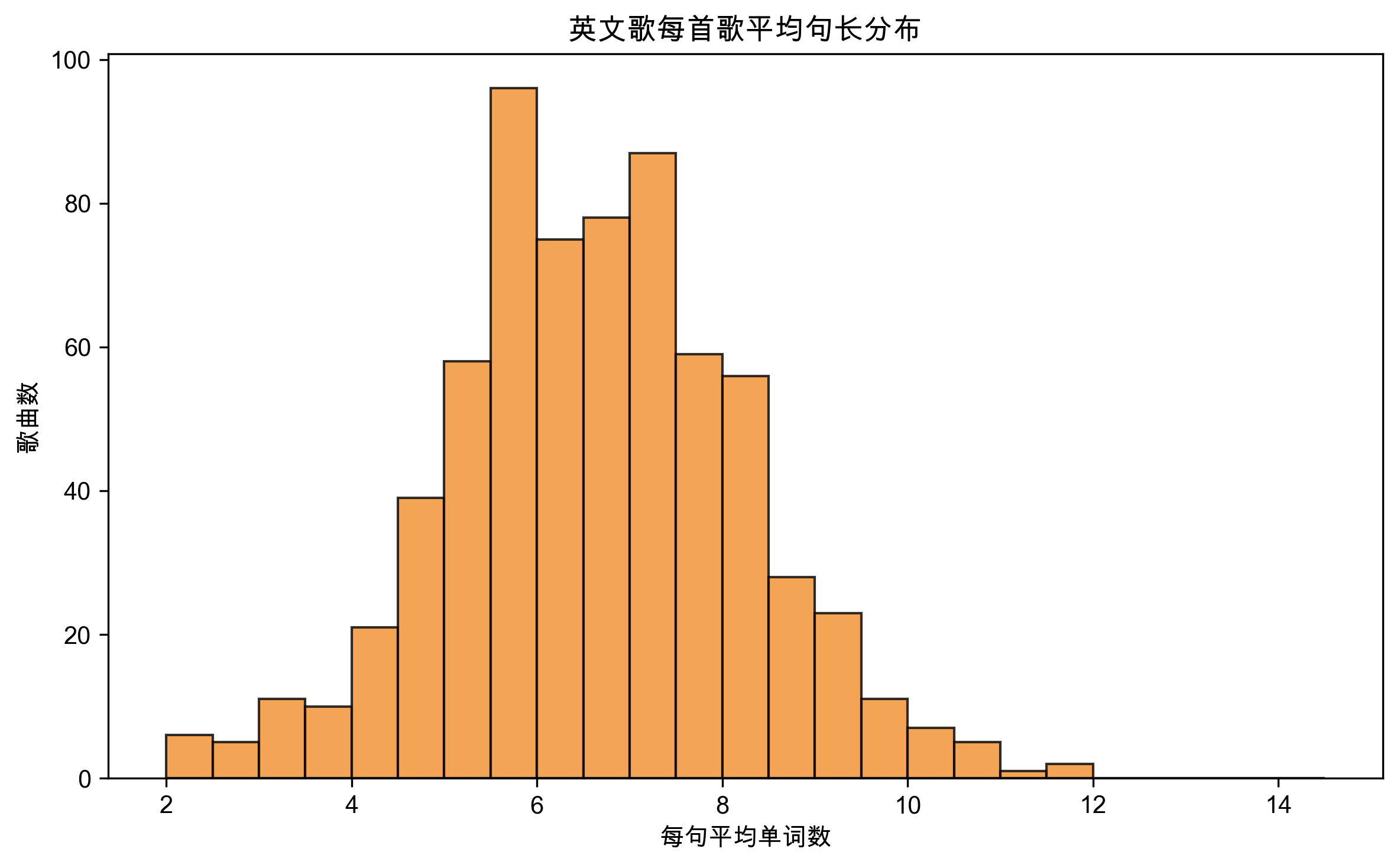
1. 歌词词频统计：乐观与美好是音乐的主流

上图为对歌曲库2020首歌曲的歌词词频统计生成的云图，代码实现如下：

可见，在剔除STOPWORDS（包括中英文中各种不表示实际意义的虚词）后，英文歌曲中出现频率最高的词是love, know, now, got, feel, see, make, tell，其中love和know二词占据绝对主流；中文歌曲上榜的词则为“世界”“未来”“明日”“今日”“希望”。可见，无论是“爱”“希望”与“明日”这些无比美好的存在，还是“知晓”“感受”“看”“做”“告诉”这般一个个打通人与人之间情感隔阂的动作，都充斥着一种对更好的未来的希冀与盼望。无论当下如何辛酸、如何苦涩，歌手们总是在提醒我们不要忘记“明天会更好”。

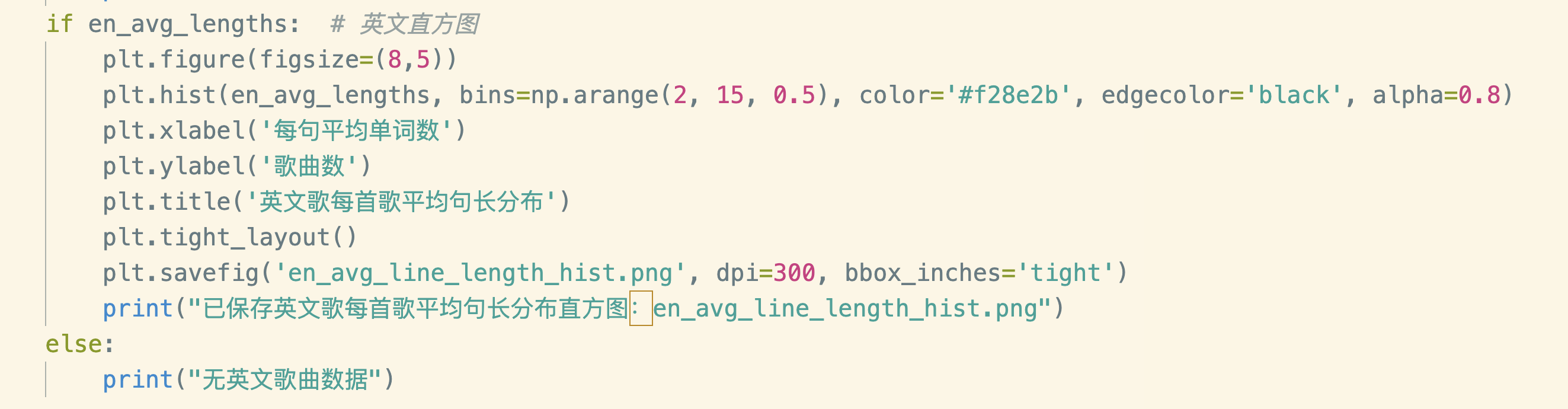
1. 歌词句长统计：短小精悍才能朗朗上口





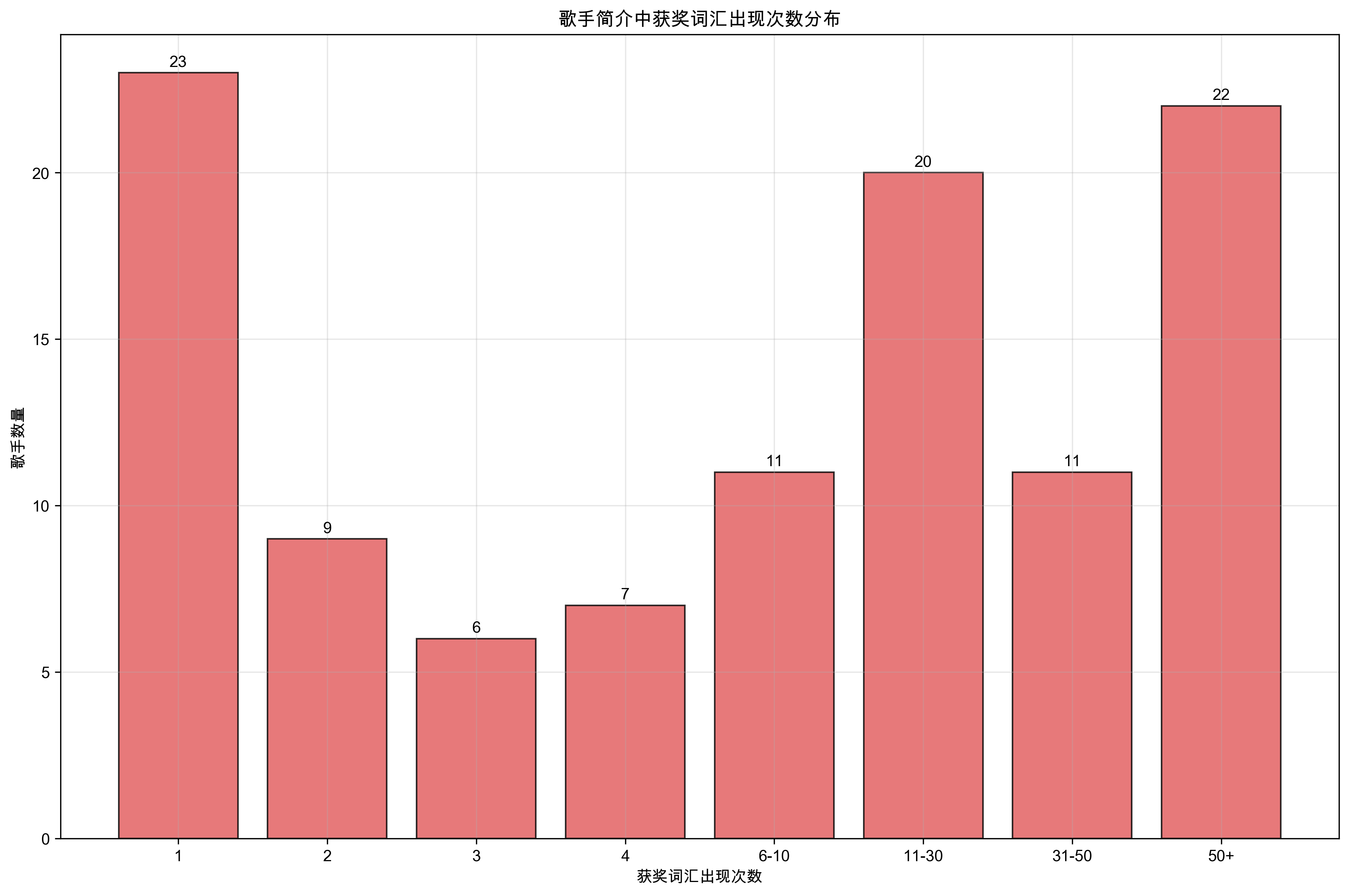
上两图分别为对所有中文和英文歌曲每首歌的平均句子长度的统计。句子以换行为分割，统计的句子长度只包括汉字或者单词的数量，不包括标点符号。下图为代码实现。





可见，90%的中文歌一句话都不超过10个字，80%的英文歌一句话都不超过8个单词。这与日常的表达习惯有明显差距——日常说话的两次换气之间通常不止说10个字，英文亦是如此，如果考虑书面语则更是天差地别。由此可知，歌词写作最重要的还是一种节奏的把控：将一句话尽可能用简短的方式表达出来，才能配合音乐对听众形成更强的冲击力。

1. 获奖难！



考虑到网易云音乐歌手简介通常由专人撰写，会详细列出歌手获奖记录，现将与获奖（中英文）相关的词汇进行了统计。代码如下：



可见，约250名歌手中超过一半从未获奖过，获奖在5次以下的更是占了大多数；但也不乏少数获奖超过50次的歌手，这些往往都是家喻户晓的国际巨星。总体来看，获奖还是呈现出相当程度的“二八定律”和“长尾效应”。获奖很难，音乐梦的道路上任重道远！