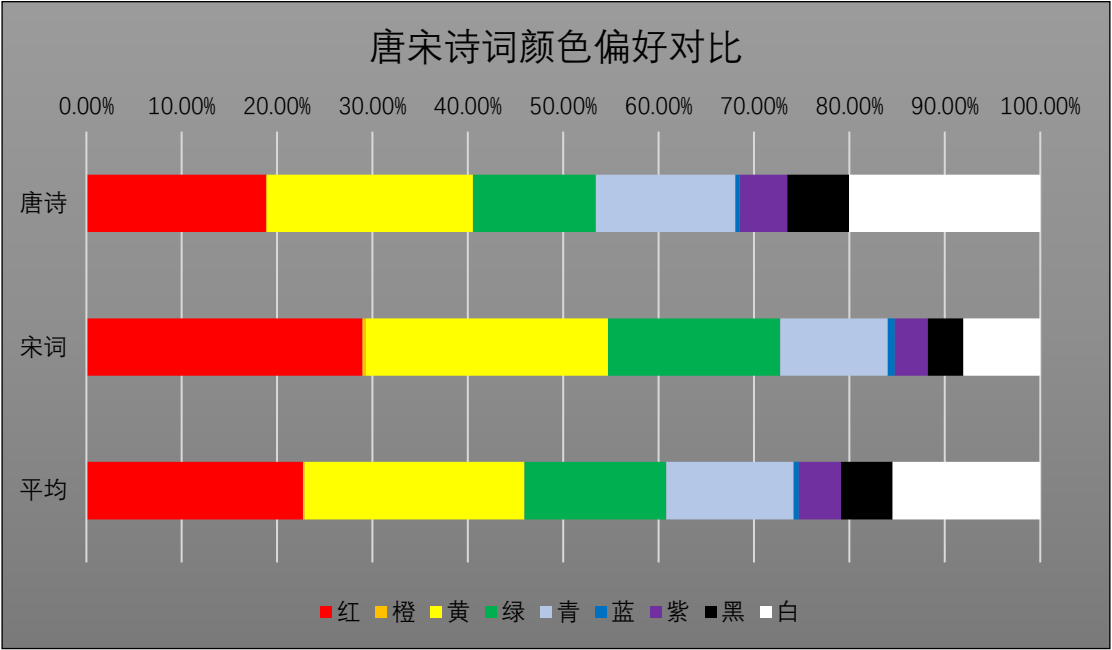


数据库原理及其应用第六周小组作业

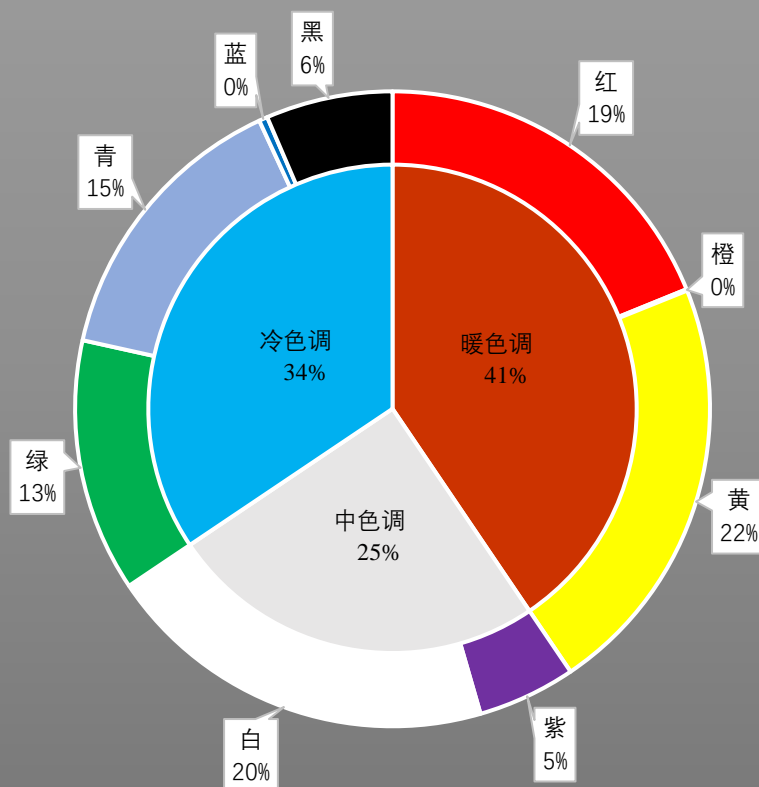
一、小组成员信息

小组组号：2
小组成员：税家晖 41934063
 刘彦升 41934049
 赵宏泽 41933008

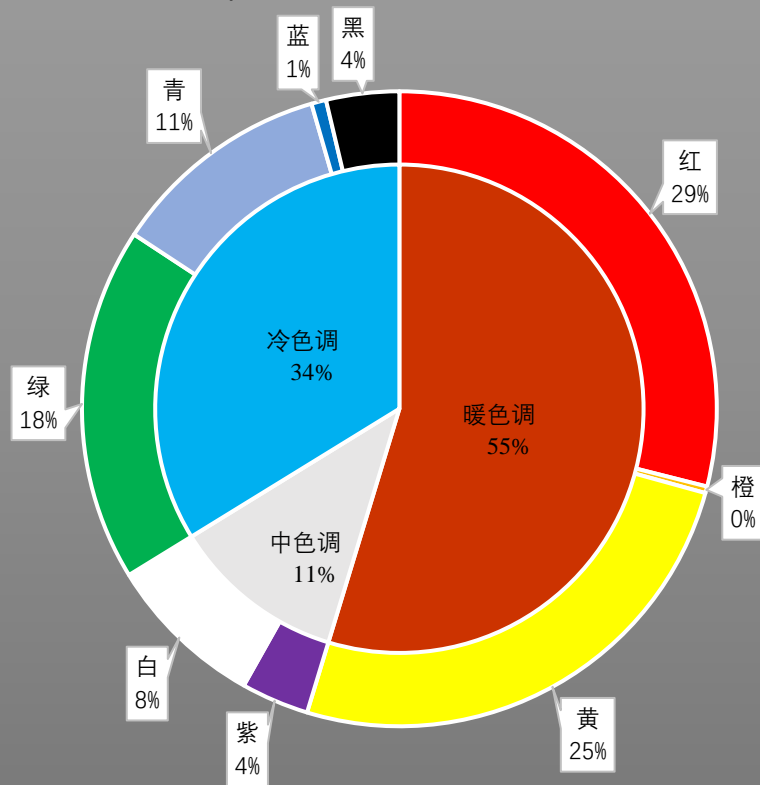
二、可视化结果

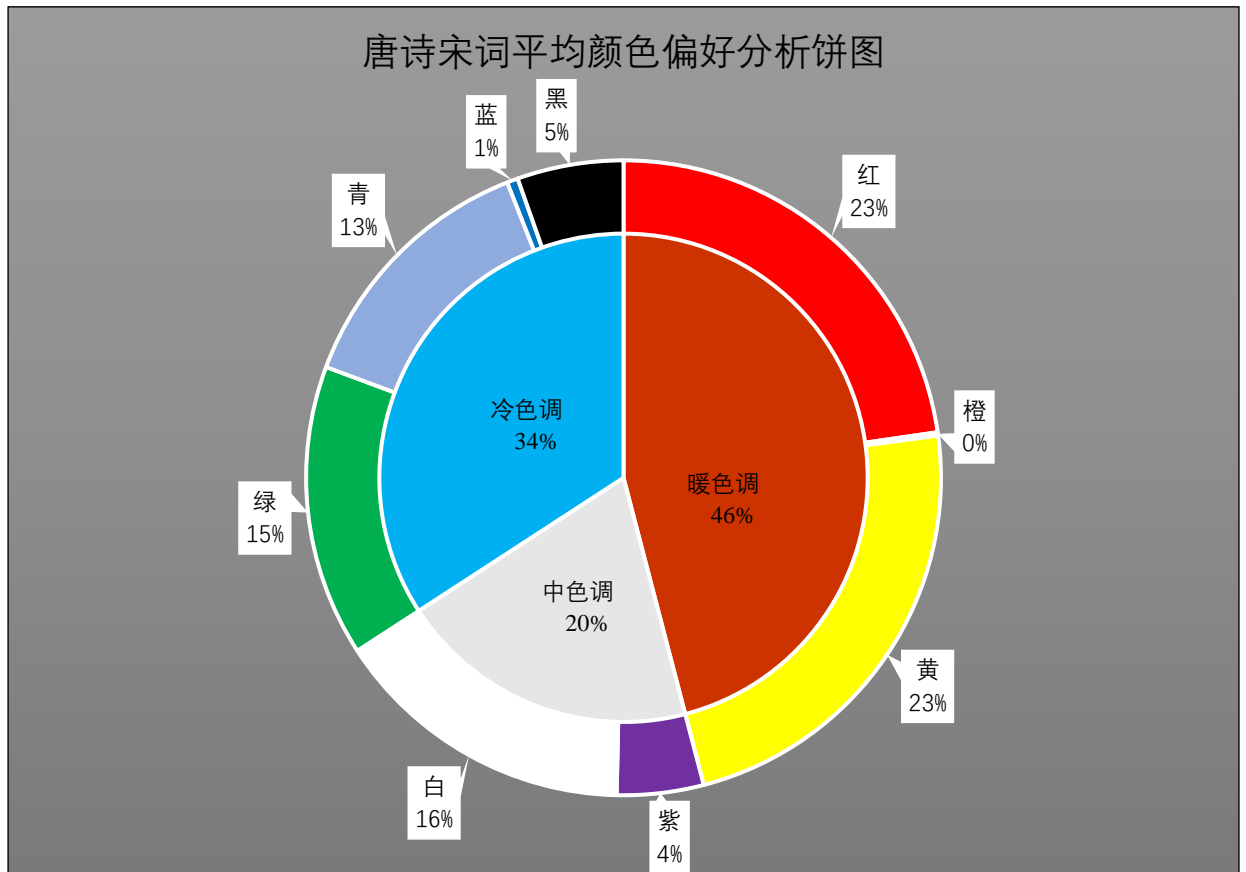


唐诗颜色偏好分析饼图



宋词颜色偏好分析饼图





从上面的几幅图中，我们可以看出，唐诗的中色调要多于宋词；而唐诗的暖色调要少于宋词。二者在冷色调上的占比几乎一致。

暖色调中，宋词在红色上较唐诗有着极为突出的表现。我们可以推测，这与宋词的婉约风格有关。

在中色调中，唐诗更为偏爱白色。白色常渲染出一种韶华易逝、悲凉的气氛，常见的有白发，白云等。

在冷色调中，唐诗与宋词的比例大致一致。但略有不同，宋词绿色和青色的占比较高，黑色则稍少。

三、附录（完成可视化过程）

第一步，我们先要统计所有的数据数目：

唐诗：总数据条数为 42862 条，非空数据条数为 42861 条

宋词：总数据条数为 20692 条，非空数据条数为 20692 条

然后创建 top100 作者的视图

```
create view tang_100
as
select top 100 author, count(*) cnt
from tangshi
where author is not null
group by author
order by cnt desc
```

宋词同理。

第二步：我们对已有的颜色分为9类，分别为红(红，赤，丹，绯，绛，朱，赭)，橙(橙)，黄(黄，金)，绿(绿，碧)，青(青，靛)，蓝(蓝)，紫(紫)，黑(黑，墨，玄，乌)，白(白)。其风格分为三类，暖色调(1)，中色调(2)，冷色调(3)。注意此处 keywords 的设置较上课时有新增。

	words	color	style
▶	红	1	1
	赤	1	1
	丹	1	1
	绯	1	1
	绛	1	1
	朱	1	1
	赭	1	1
	橙	2	1
	黄	3	1
	金	3	1
	白	9	2
	绿	4	3
	碧	4	3
	青	5	3
	靛	5	3
	蓝	6	3
	紫	7	2
	黑	8	3
	墨	8	3
	玄	8	3
	乌	8	3

对唐诗进行分析，首先我们先按照颜色，筛选出前 100 名诗人的数据

```
select color, count(*) cnt
from tangshi a, keywords b, tang_100 c
where patindex('%'+rtrim(words)+'%', contents)>0
and a.author = c.author
group by color
order by cnt desc
```

再按照风格进行筛选

```
select style, count(*) cnt
from tangshi a, keywords b, tang_100 c
where patindex('%'+rtrim(words)+'%', contents)>0
and a.author=c.author
group by style
order by cnt desc
```

对于宋词我们做同样的处理，代码为

```
select color, count(*) cnt
from songci a, keywords b, song_100 c
where patindex('%'+rtrim(words)+'%', contents)>0
and a.author = c.author
group by color
order by cnt desc
```

```
select style, count(*) cnt
from songci a, keywords b, song_100 c
where patindex('%'+rtrim(words)+'%', contents)>0
and a.author=c.author
group by style
order by cnt desc
```

同时我们发现，有不少的词句中重复运用了颜色。

序号	词人	词牌名	词内容
1	晏殊	渔家傲	杨柳风前香百步。盘心碎点真珠露。疑是水仙开洞府。妆景趣。红幢绿盖朝天路。小鸭飞来稠闹处。三三两两...
2	晏殊	渔家傲	叶下鸂鶒眠未稳。风翻露飏香成阵。仙女出游知远近。羞借问。饶将绿扇遮红粉。一搦蕊黄沾雨润。天人乞与...
3	晏殊	浣溪沙	沙上寒鸥接翼飞。潮生潮落水东西。征船鸣橹趁潮归。望断碧云无锦字，漫题红叶有新诗。黄昏微雨倚阑时。
4	晏殊	点绛唇	香雪飘零，暖风着柳笼丝雨。恼人情绪。春事还如许。宝勒朱轮，共结寻芳侣。东郊路。乱红深处。醉拍黄金缕。
5	晏殊	长相思	村姑儿。红袖衣。初发黄梅插稻时。双双女伴随。长歌诗。短歌诗。歌里真情恨别离。休言伊不知。
6	晏殊	七娘子	天涯触目伤离绪。登临况值秋光暮。手拈黄花，凭谁分付。雍雍雁落蒹葭浦。凭高目断桃溪路。屏山楼外青无...
7	晏殊	采桑子慢	愁春未醒，还是清和天气。对浓绿阴中庭院，燕语莺啼。数点新荷翠钿，轻泛水平池。一帘风絮，才晴又雨，...
8	晏殊	金明池	帐锦笼庭，囊香飘榭，过了芳时强半。觅残红、蜂须趁日，占新绿、莺喉咤暖。数花期、望得春来，春去也、...
9	晏殊	临江仙	多病渊明刚止酒，不禁秋蕊浮香。饮船歌板已兼忘。吴霜羞鬓改，无语对红妆。小拈青枝撩鼻观，绝胜娇额涂...
10	晏殊	十样花	陌上风光浓处。第一寒梅先吐。待得春来也，香销减，态凝伫。百花休漫妒。陌上风光浓处。繁杏枝头春聚。...
11	晏殊	念奴娇	江天雨霁，正露荷擎翠，风槐摇绿。试问秦楼今夜里，愁到阑干几曲。笑拈黄花，重题红叶，无奈归期促。暮...
12	晏殊	念奴娇	蕊香深处，逢上巳、生怕花飞红雨。万点胭脂遮翠袖，谁识黄昏凝伫。烧烛呈妆，传杯绕槛，莫放春归去。垂...
13	晏殊	忆王孙	小院春登。晴窗霞透。把雨燕脂，倚风翠袖。芳意恼乱人多。暖金荷。多情不分群葩后。伤春瘦。浅黛眉尖秀...

查询同时含有红黄的词可以用如下的代码：

```
select * from songci
where patindex('%红%',contents)>0 and patindex('%黄%',contents)>0
```

第三步：利用 EXCEL 等软件对上述数据分析，总的统计结果如下所示

序号	颜色	唐诗	宋词	总计值	唐诗	宋词	平均	风格	风格	唐诗	宋词	平均	唐诗	宋词	平均
1	红	5251	4977	10228	18.85%	28.94%	22.70%	暖色调	暖色调	11287	9406	20693	40.52%	54.70%	45.93%
2	橙	21	58	79	0.08%	0.34%	0.18%	暖色调	中色调	6980	1982	8962	25.06%	11.53%	19.89%
3	黄	6015	4371	10386	21.60%	25.42%	23.05%	暖色调	冷色调	9586	5808	15394	34.42%	33.78%	34.17%
4	绿	3589	3100	6689	12.89%	18.03%	14.85%	冷色调		27853	17196	45049			
5	青	4077	1940	6017	14.64%	11.28%	13.36%	冷色调							
6	蓝	121	129	250	0.43%	0.75%	0.55%	冷色调							
7	紫	1394	594	1988	5.00%	3.45%	4.41%	中色调							
8	黑	1799	639	2438	6.46%	3.72%	5.41%	冷色调							
9	白	5586	1388	6974	20.06%	8.07%	15.48%	中色调							
	总计	27853	17196	45049			100.00%								

最后进行可视化，得到需要的图表。