Data Scientist- Stackroute Labs- Offline Assignment

Problem Statement - Your client in an NGO who takes initiatives to improve primary education in India. They want to carry out this program in Karnataka. It wants to target districts that fall behind in areas such as education infrastructure, education awareness, and demographic features. Identify such districts that could be targeted in its first phase

Note – The entire R Code has been attached in a Jupyter file.

Submitted by Vicky Crasto | vickycrasto@gmail.com | 8123363096

Data Exploration and variable creation

Dataset

The data regarding the primary education district wise(2015- 2016) has been downloaded from http://schoolreportcards.in/SRC-New/Links/DISEPublication.aspx
Census data- 2011 has been downloaded from http://www.census2011.co.in/

Data Exploration, creation of additional Variables

After sub setting the data for the state of Karnataka it was imported into R.
The data was systematic and did not have any missing values, however a data dictionary was created to understand the variable name.

Overview of the dataset

- **Dataset 1** gave information about the distribution of the schools in each district and provide statistics about the availability of the basic infra structure in the school
- **Dataset 2** gave information about the student enrollment in each grade and the quality of the teacher.
- **Dataset 3** gave information about the pervious year's data regard the distribution of the schools in each district and the availability of the basic infra structure in the school
- **Dataset4** was the census data set with all the details collected during the census.

Variable creation

Using dataset 1 the following variable were calculated

- Percentage of children in the age group of 6 to 15
- Percentage of single classroom school
- Percentage of single teacher school
- Percentage of Schools Approachable by All Weather Road
- Percentage of schools with Playground Facility
- Percentage of Schools with Boundary wall
- Percentage of Schools with Girls' Toilet
- Percentage of Schools with Boys' Toilet
- Percentage of Schools with Drinking Water
- Percentage of Schools with Electricity
- Percentage of schools with computers
- Percentage of schools with Schools with Ramp (where needed)
- Percentage of schools providing mid meal (govt aid)
- Percentage of Schools with Kitchen-Shed (Government & Aided Schools)
- Percentage of Schools where Mid-Day Meal is Provided and Prepared in School Premises
- Ratio of teacher to student

Variable creation

Using dataset 2 the following variable were calculated

- Percentage of enrollment for boys grade 1 to 5
- Percentage of enrollment for boys grade 6 to 9
- Percentage of enrollment for girls grade 1 to 5
- Percentage of enrollment for girls grade 6 to 9
- Percentage of enrollment for special need boys grade 1 to 5
- Percentage of enrollment for special need boys grade 6 to 8
- Percentage of enrollment for special need girls grade 1 to 5
- Percentage of enrollment for special need girls grade 6 to 9
- Percentage of enrollment of SC girls in primary
- Percentage of enrollment of SC girls in upper primary
- Percentage of enrollment of ST girls in primary
- Percentage of enrollment of ST girls in upper primary
- Percentage of enrollment of Muslim girls in primary
- Percentage of enrollment of Muslim girls in upper primary
- Percentage of class in good condition
- Percentage of classes needing minor repair
- Percentage of classes needing major repair

Using Dataset 4 the following variable have been calculated

- Percentage of literate
- Percentage of male literate
- Percentage of female literate
- Percentage of SC population
- Percentage of ST population
- Percentage of Muslim population
- Percentage of urban household
- Percentage of rural household
- Percentage of household with electricity

Scorecard Model

In order to able to identify the district to be focused, relevant variables are selected and divided into 3 broad heading and each variable is given score.

All the variable are given equal weights.

The total score in each heading is calculated and the basis the overall score the districts to focused are identified.

Each of the relevant variable has been expressed as percentage, this is converted to a score out of 1 marks by dividing by 100.

For variable with a desired quality or positive aspect related to primary education has a positive mark but an undesired quality is expressed as a negative mark for the state.

The 3 board headings are

- **School infrastructure**
- **Focus on girl child education**
- **Overall quality of education and demographic feature**

The below table shows the district wise break up for group and the variable in each group

School infrastructure

Variable under this group are

- Approach road to the school
- Playground
- Boundary wall
- Girl's Toilet
- Boy's Toilet
- Drinking water
- Electricity
- Computers
- Ramp
- Kitchen shed and Mid day meal
- Classroom Condition

Rank	District	Approach_road	playground	boundarywall	girls_toilet	boys_toilet	drinking_water	electricity	computer	ramp	kitchen_shed	classroom_cond	sch_infr_total_score
1	Yadgir	0.83	0.55	0.54	0.99	0.96	1.00	0.91	0.20	0.77	0.54	0.30	7.58
2	Gulbarga	0.90	0.61	0.62	1.00	0.99	1.00	0.93	0.30	0.70	0.42	0.31	7.77
3	Bellary	0.89	0.59	0.73	0.99	0.99	1.00	0.97	0.37	0.70	0.36	0.31	7.90
4	Dharwad	0.93	0.74	0.79	0.96	0.95	1.00	0.96	0.55	0.72	0.00	0.31	7.90
5	Bijapur	0.91	0.76	0.56	0.97	0.96	1.00	0.93	0.32	0.73	0.54	0.31	8.00
6	Bangalore	0.95	0.69	0.88	1.00	0.99	1.00	1.00	0.76	0.38	0.04	0.30	8.00
7	Bidar	0.91	0.70	0.67	1.00	0.99	1.00	0.91	0.23	0.71	0.60	0.32	8.03
8	Raichur	0.80	0.61	0.68	1.00	0.99	1.00	0.98	0.24	0.76	0.67	0.31	8.04
9	Belgaum	0.93	0.77	0.72	0.98	0.97	1.00	0.90	0.30	0.81	0.47	0.32	8.18
10	Koppal	0.90	0.66	0.80	0.99	0.98	1.00	0.93	0.27	0.76	0.62	0.30	8.22
11	Chamrajnagar	0.87	0.52	0.81	1.00	0.97	1.00	0.94	0.30	0.85	0.67	0.32	8.25
12	Chikballapura	0.90	0.52	0.75	0.99	0.99	1.00	0.97	0.25	0.86	0.75	0.32	8.30
13	Kolar	0.94	0.50	0.78	0.96	0.91	1.00	0.97	0.35	0.86	0.82	0.31	8.40
14	Ramanagara	0.94	0.47	0.75	1.00	1.00	1.00	1.00	0.28	0.87	0.79	0.31	8.40
15	Tumkur	0.91	0.49	0.81	1.00	1.00	1.00	0.98	0.21	0.89	0.82	0.31	8.41
16	Davanagere	0.93	0.63	0.81	0.99	0.98	1.00	0.96	0.31	0.78	0.72	0.30	8.42
17	Chitradurga	0.93	0.58	0.75	1.00	0.99	1.00	0.99	0.20	0.88	0.80	0.31	8.43
18	Shimoga	0.90	0.63	0.85	1.00	1.00	1.00	0.97	0.18	0.84	0.75	0.31	8.44
19	Bagalkot	0.94	0.72	0.80	0.98	0.97	1.00	0.98	0.31	0.78	0.66	0.31	8.44
20	Hassan	0.99	0.56	0.74	1.00	0.99	1.00	1.00	0.20	0.88	0.81	0.32	8.49
21	Bangalore Rural	0.95	0.52	0.87	1.00	1.00	1.00	0.99	0.31	0.83	0.72	0.32	8.50
22	Uttara Kannada	0.84	0.70	0.76	1.00	0.99	1.00	0.99	0.28	0.93	0.80	0.32	8.60
23	Mandya	0.97	0.59	0.84	1.00	0.99	1.00	1.00	0.27	0.85	0.77	0.32	8.60
24	Gadag	0.95	0.79	0.87	0.94	0.92	1.00	0.99	0.47	0.79	0.61	0.31	8.67
25	Mysore	0.97	0.61	0.91	1.00	0.98	1.00	0.98	0.41	0.79	0.70	0.32	8.67
26	Chikmagalur	0.93	0.71	0.83	1.00	1.00	1.00	0.99	0.27	0.87	0.77	0.32	8.67
27	Udupi	0.94	0.83	0.68	1.00	0.98	1.00	0.99	0.45	0.85	0.77	0.32	8.80
28	Haveri	0.99	0.65	0.78	0.97	0.96	1.00	0.99	0.70	0.83	0.67	0.31	8.84
29	Dakshina Kannada	0.96	0.94	0.83	1.00	0.99	1.00	1.00	0.48	0.79	0.70	0.31	8.99
30	Kodagu	0.98	0.84	0.90	1.00	1.00	1.00	0.98	0.54	0.79	0.72	0.30	9.04

Focus on girl child education

Variable under this group are

- Overall enrollment (girls)
- Primary level enrollment
- Upper primary enrollment
- Special need – Primary
- Special needs – Upper Primary
- SC primary enrollment
- SC upper primary enrollment
- ST primary enrollment
- ST upper primary enrollment
- Muslim primary enrollment
- Muslim upper primary enrollment

Rank	District	overall_enrollment	primary_enrollment	upper_prim_enroll	spec_prim_enroll	spec_upper_prim_e	SC_primary	SC_upper_pr	ST_primary	ST_upper_pr	muslim_prim	muslim_upper_prim	girl_edu_tot
1	Yadgir	0.46	0.31	0.16	0.0029	0.0010	0.48	0.43	0.48	0.45	0.48	0.50	3.74
2	Bellary	0.48	0.29	0.19	0.0030	0.0015	0.49	0.47	0.49	0.48	0.48	0.47	3.84
3	Bijapur	0.50	0.30	0.18	0.0029	0.0013	0.48	0.46	0.48	0.47	0.48	0.49	3.84
4	Gulbarga	0.51	0.30	0.18	0.0032	0.0019	0.49	0.46	0.46	0.47	0.49	0.51	3.87
5	Tumkur	0.41	0.27	0.21	0.0030	0.0017	0.49	0.48	0.50	0.47	0.52	0.53	3.89
6	Raichur	0.46	0.29	0.18	0.0029	0.0012	0.48	0.46	0.48	0.47	0.53	0.53	3.90
7	Bangalore	0.47	0.29	0.19	0.0012	0.0006	0.49	0.49	0.48	0.48	0.50	0.52	3.91
8	Ramanagara	0.45	0.27	0.21	0.0041	0.0032	0.49	0.48	0.49	0.47	0.52	0.53	3.92
9	Mandya	0.43	0.26	0.21	0.0032	0.0026	0.49	0.49	0.47	0.48	0.56	0.54	3.93
10	Chitradurga	0.46	0.27	0.21	0.0043	0.0025	0.49	0.49	0.48	0.49	0.52	0.52	3.93
11	Uttara Kannad	0.45	0.26	0.21	0.0066	0.0042	0.48	0.47	0.51	0.46	0.55	0.55	3.94
12	Kolar	0.45	0.27	0.21	0.0021	0.0011	0.49	0.49	0.49	0.48	0.53	0.54	3.96
13	Bagalkot	0.48	0.28	0.19	0.0032	0.0014	0.49	0.48	0.49	0.50	0.52	0.53	3.96
14	Bidar	0.53	0.30	0.19	0.0047	0.0015	0.49	0.48	0.49	0.50	0.49	0.51	3.97
15	Koppal	0.47	0.29	0.19	0.0031	0.0018	0.49	0.47	0.49	0.49	0.53	0.55	3.98
16	Chikballapura	0.46	0.27	0.21	0.0034	0.0023	0.49	0.49	0.50	0.48	0.53	0.54	3.98
17	Chamrajnagar	0.45	0.27	0.21	0.0031	0.0017	0.48	0.49	0.49	0.48	0.54	0.57	3.98
18	Dharwad	0.51	0.28	0.21	0.0030	0.0019	0.49	0.48	0.48	0.49	0.53	0.54	3.99
19	Gadag	0.49	0.28	0.20	0.0048	0.0037	0.49	0.48	0.49	0.49	0.54	0.55	4.01
20	Davanagere	0.49	0.27	0.21	0.0030	0.0020	0.49	0.49	0.49	0.49	0.54	0.55	4.01
21	Mysore	0.48	0.27	0.21	0.0026	0.0020	0.49	0.49	0.49	0.50	0.53	0.54	4.02
22	Belgaum	0.56	0.28	0.20	0.0027	0.0014	0.49	0.48	0.49	0.48	0.53	0.53	4.05
23	Shimoga	0.49	0.26	0.22	0.0042	0.0026	0.49	0.49	0.48	0.49	0.57	0.57	4.06
24	Bangalore Rur	0.48	0.28	0.20	0.0027	0.0020	0.49	0.49	0.48	0.47	0.62	0.55	4.06
25	Chikmagalur	0.47	0.26	0.22	0.0034	0.0023	0.49	0.50	0.49	0.50	0.57	0.57	4.07
26	Udupi	0.50	0.26	0.22	0.0041	0.0025	0.49	0.47	0.49	0.47	0.59	0.59	4.09
27	Haveri	0.47	0.27	0.21	0.0037	0.0026	0.48	0.49	0.49	0.49	0.59	0.60	4.10
28	Dakshina Kann	0.50	0.26	0.22	0.0030	0.0022	0.48	0.49	0.49	0.49	0.59	0.59	4.11
29	Hassan	0.46	0.27	0.22	0.0031	0.0021	0.50	0.50	0.51	0.52	0.57	0.58	4.13
30	Kodagu	0.48	0.26	0.22	0.0030	0.0023	0.50	0.50	0.51	0.50	0.61	0.59	4.17

Overall quality of education and demographic feature

Variable under this group are

- Single class school
- Single teacher school
- Pupil teacher ratio
- Literate population
- Male literate population
- Female literate population

A negative mark indicate that it is undesired feature in the school and should be focused on removing it completely.
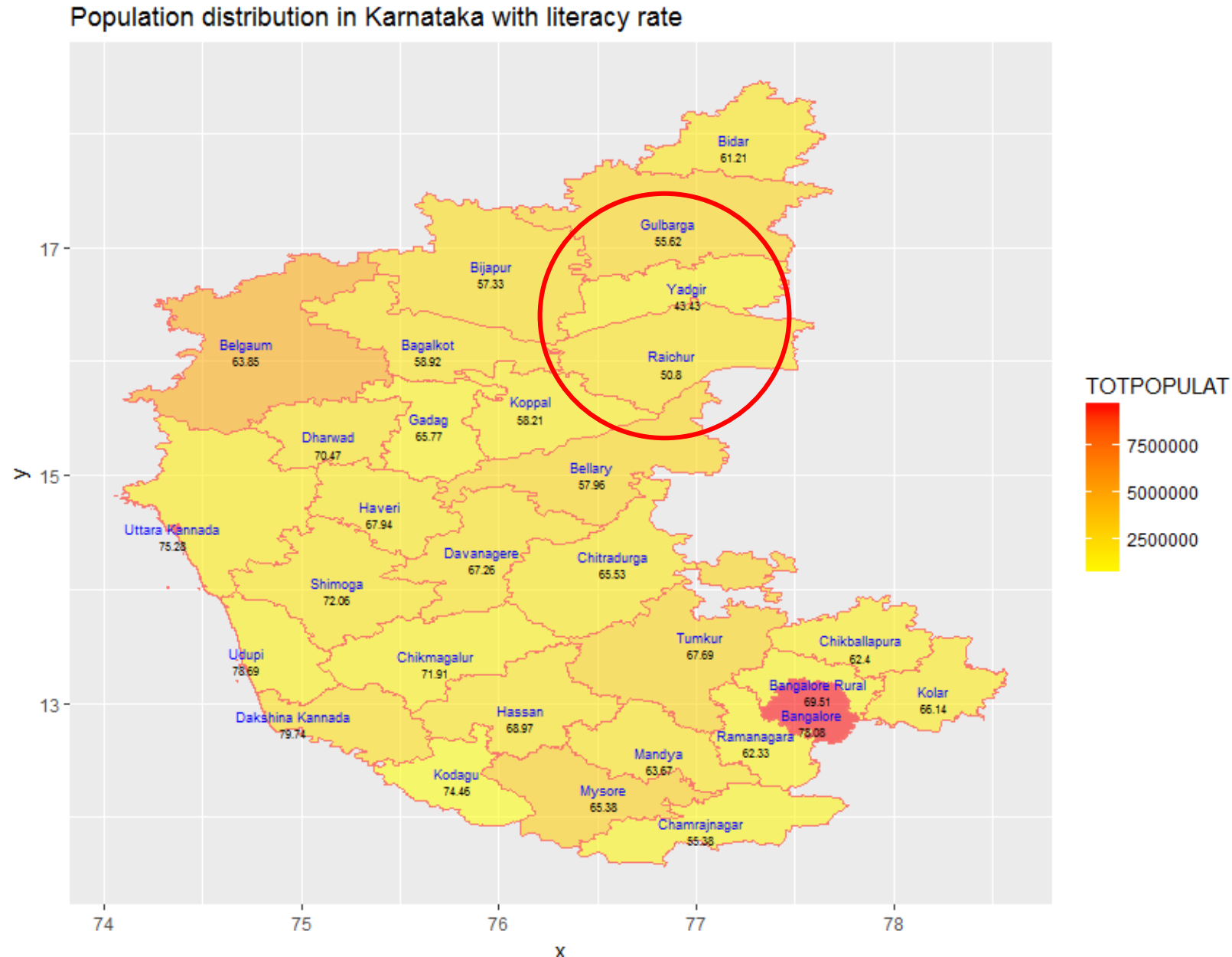
As per standard set by Indian government, the pupil teacher ratio should be 30:1, for 30 students there must be one teacher. For district maintaining this standard have score or 1.

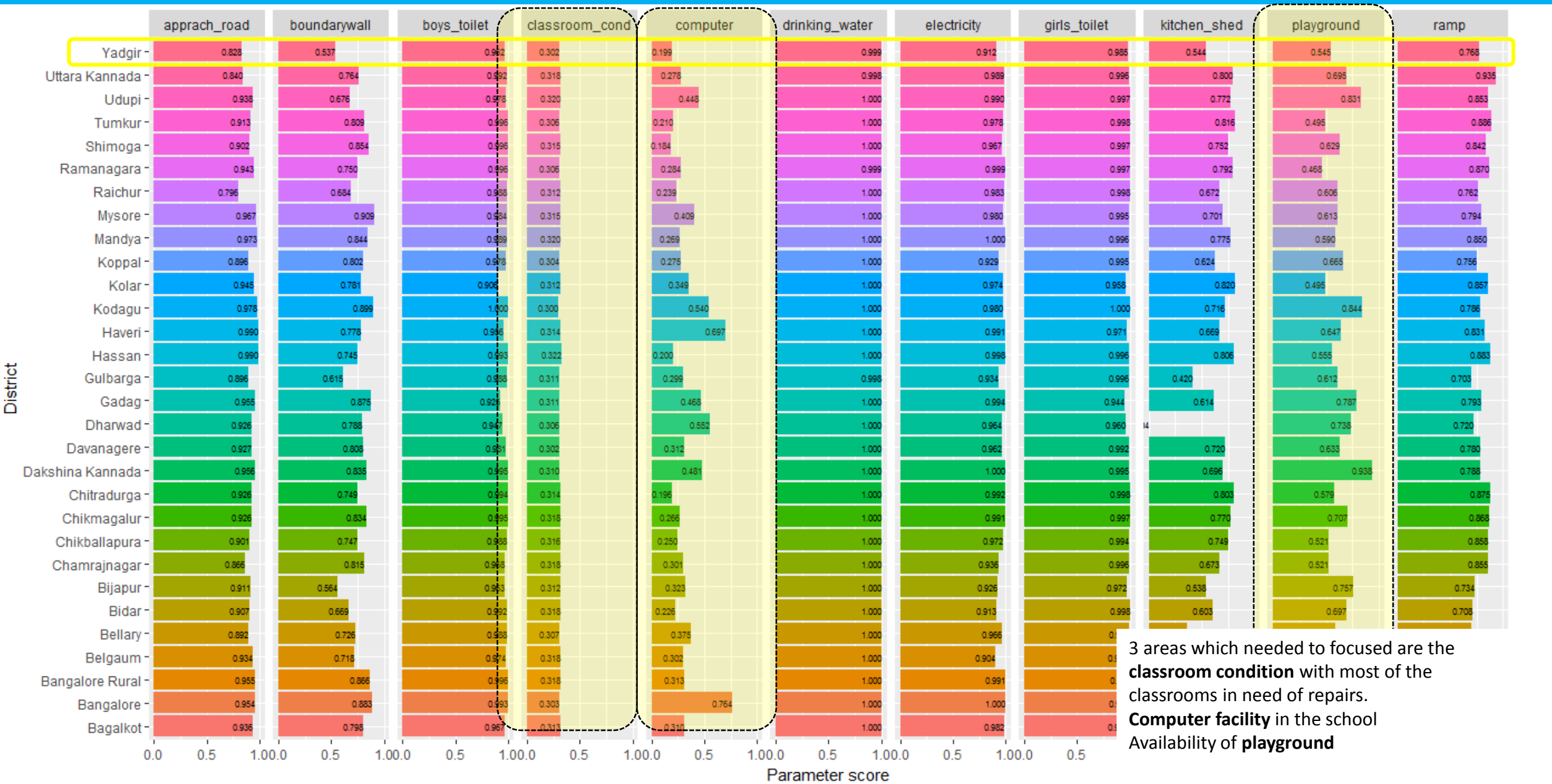| Sr.No.	District	Single_class	single_teacher	pupil_teacher_ratio	literate_pop	male_literate_pop	female_literate_pop	educ_qua_total_score
1	Yadgir	-0.0119	-0.0901	0.5	0.43	0.26	0.17	1.27
2	Bellary	-0.0133	-0.0617	0.75	0.58	0.33	0.25	1.83
3	Bagalkot	-0.0027	-0.0538	0.75	0.59	0.34	0.25	1.87
4	Raichur	-0.0194	-0.0899	1	0.51	0.30	0.21	1.91
5	Belgaum	-0.0279	-0.0817	0.75	0.64	0.36	0.28	1.92
6	Chamrajna	-0.0422	-0.1256	1	0.55	0.31	0.25	1.94
7	Ramanaga	-0.1211	-0.1767	1	0.62	0.35	0.27	1.95
8	Gulbarga	-0.0138	-0.0566	1	0.56	0.32	0.23	2.04
9	Chikballapu	-0.0587	-0.1446	1	0.62	0.35	0.27	2.04
10	Mandya	-0.032	-0.1558	1	0.64	0.35	0.28	2.09
11	Bijapur	-0.0176	-0.0384	1	0.57	0.34	0.24	2.09
12	Koppal	-0.0091	-0.0564	1	0.58	0.34	0.24	2.10
13	Kolar	-0.0956	-0.1262	1	0.66	0.37	0.29	2.10
14	Bangalore	-0.1039	-0.1432	1	0.70	0.39	0.31	2.14
15	Tumkur	-0.0734	-0.13	1	0.68	0.38	0.30	2.15
16	Bidar	-0.0057	-0.0531	1	0.61	0.35	0.26	2.17
17	Hassan	-0.0798	-0.1118	1	0.69	0.38	0.31	2.19
18	Chitradurg	-0.0487	-0.0645	1	0.66	0.37	0.29	2.20
19	Mysore	-0.035	-0.0723	1	0.65	0.35	0.30	2.20
20	Dakshina K	-0.0027	-0.0939	0.75	0.80	0.41	0.38	2.25
21	Gadag	-0.0023	-0.0336	1	0.66	0.37	0.28	2.28
22	Davanager	-0.0169	-0.0384	1	0.67	0.37	0.30	2.29
23	Shimoga	-0.0342	-0.1107	1	0.72	0.39	0.34	2.30
24	Chikmagal	-0.0297	-0.1119	1	0.72	0.39	0.33	2.30
25	Haveri	-0.008	-0.0359	1	0.68	0.38	0.30	2.31
26	Uttara Kan	-0.0299	-0.1289	1	0.75	0.41	0.35	2.35
27	Kodagu	-0.0092	-0.1099	1	0.74	0.39	0.36	2.37
28	Dharwad	-0.013	-0.0261	1	0.70	0.39	0.32	2.37
29	Udupi	-0.0062	-0.1007	1	0.79	0.40	0.39	2.47
30	Bangalore	-0.0171	-0.0422	1	0.78	0.42	0.36	2.50

Population Distribution and Literacy Rate

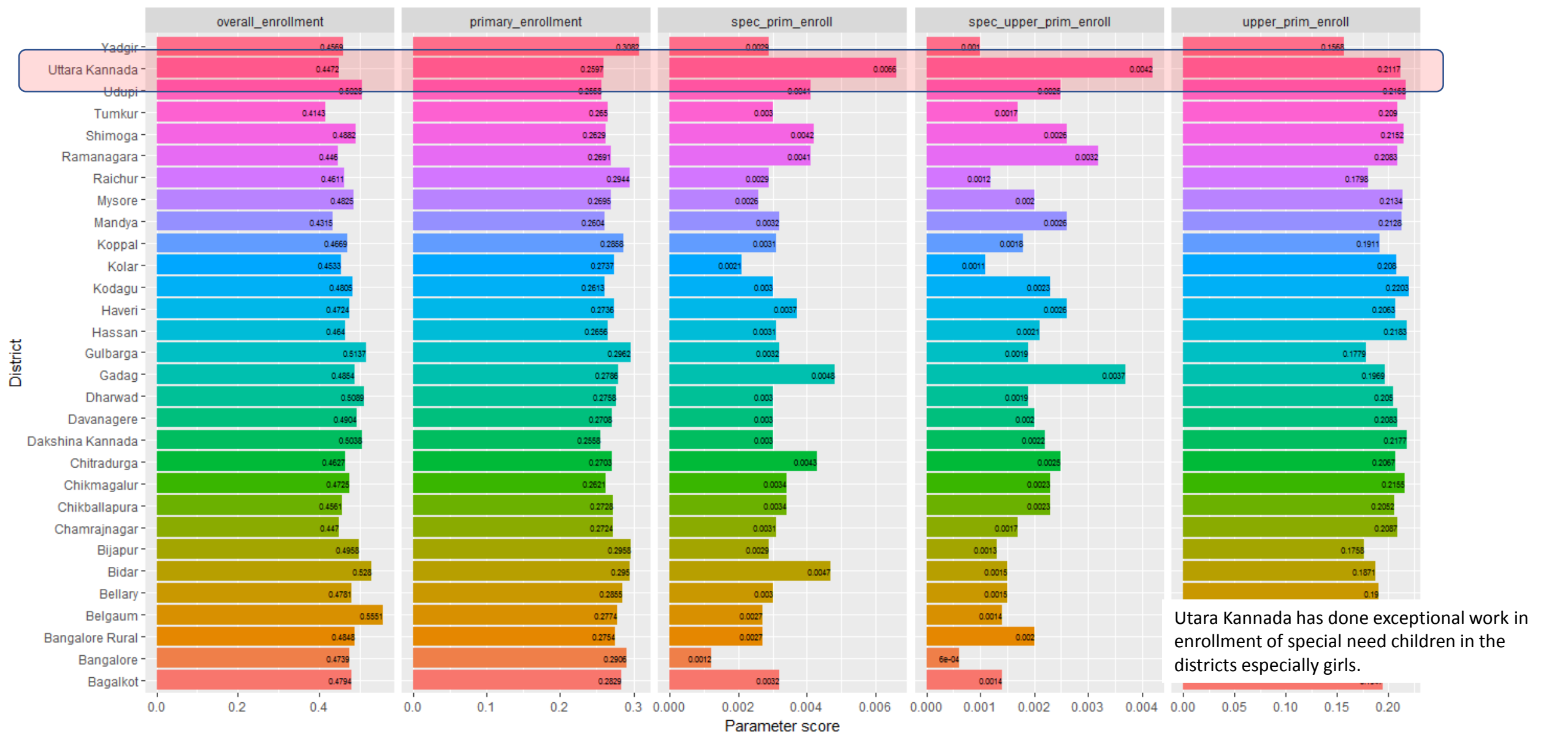As highlighted the overall literacy rate is low in Gulbarga, Yadgir and Raichur district.

A deeper investigation in these districts will help us understand the quality of primary education

Population distribution in Karnataka with literacy rate

Comparing School infrastructure Scores

3 areas which needed to focused are the **classroom condition** with most of the classrooms in need of repairs. **Computer facility** in the school Availability of **playground**

Comparing Girl Child Enrollment

Utara Kannada has done exceptional work in enrollment of special need children in the districts especially girls.

Comparing Education Quality

Ramanagara needs to focus on improving education quality

Bangalore Rural surprising has a high level of single class room which must investigated further.
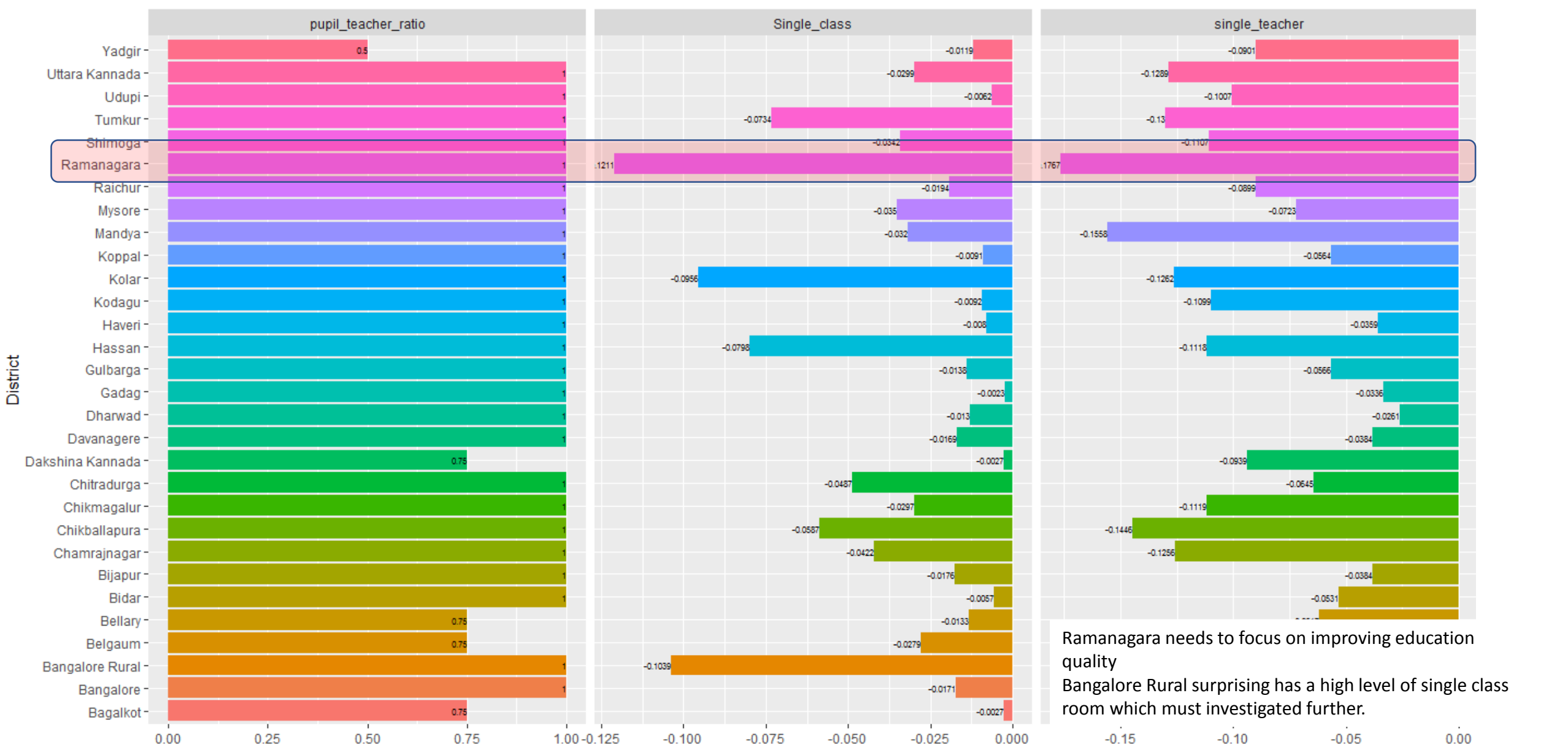
Girl Child education Focus in SC communities

Here the numbers below the district name indicate the percentage of girl child enrollment of the total SC enrollment in primary and upper primary.

Kolar has a high SC population along with Gulbarga, Yadgir and Raichur but their focus of education girl child is low
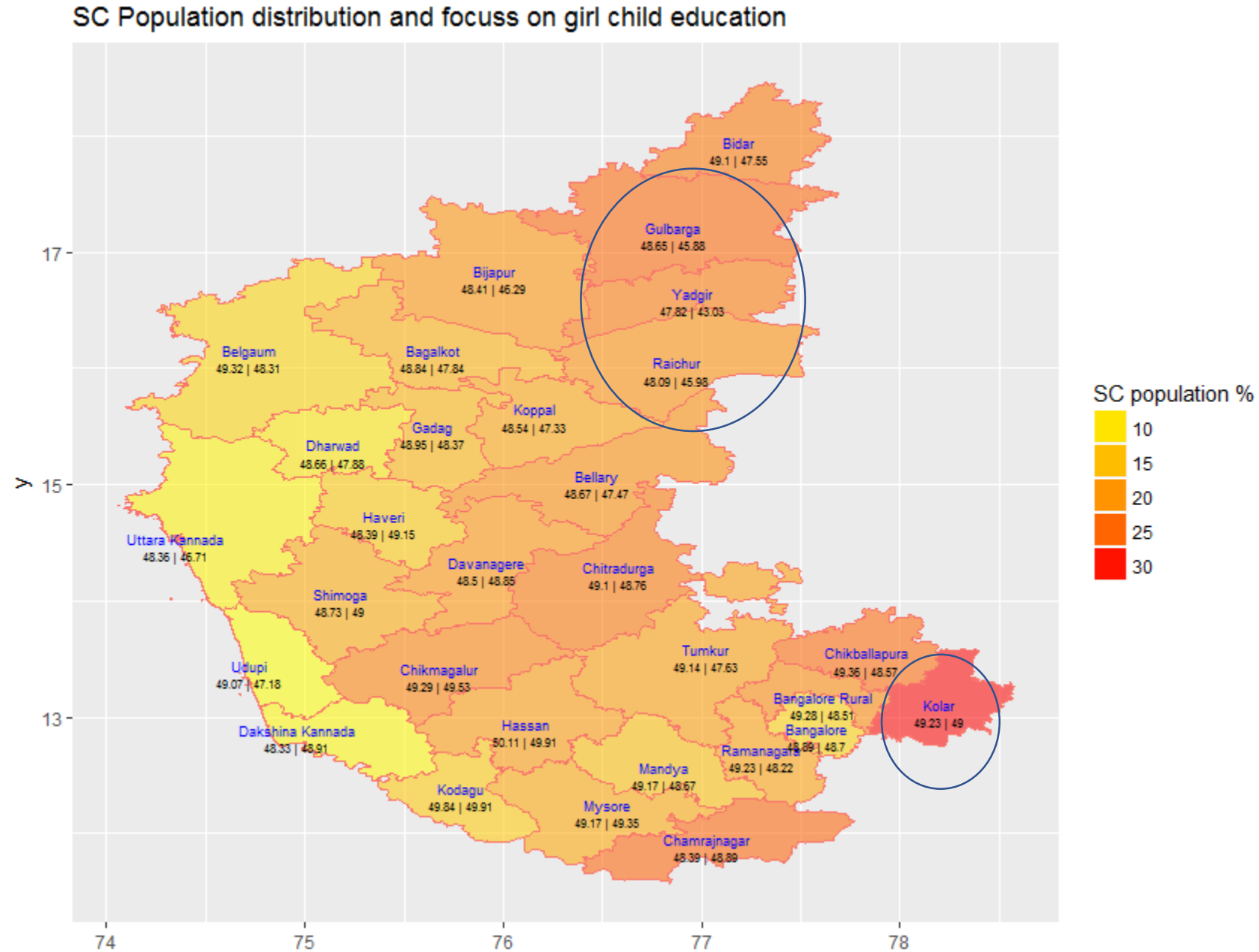

SC Population distribution and focuss on girl child education

Girl Child education Focus in ST communities

Here the numbers below the district name indicate the percentage of girl child enrollment of the total ST enrollment in primary and upper primary.

Yadgir and Raichur have a high ST population but their focus of education girl child is low

ST Population distribution and focuss on girl child education

ST population %
5
10
15

Girl Child education Focus in Muslim communities

Here the numbers below the district name indicate the percentage of girl child enrollment of the total muslim enrollment in primary and upper primary.
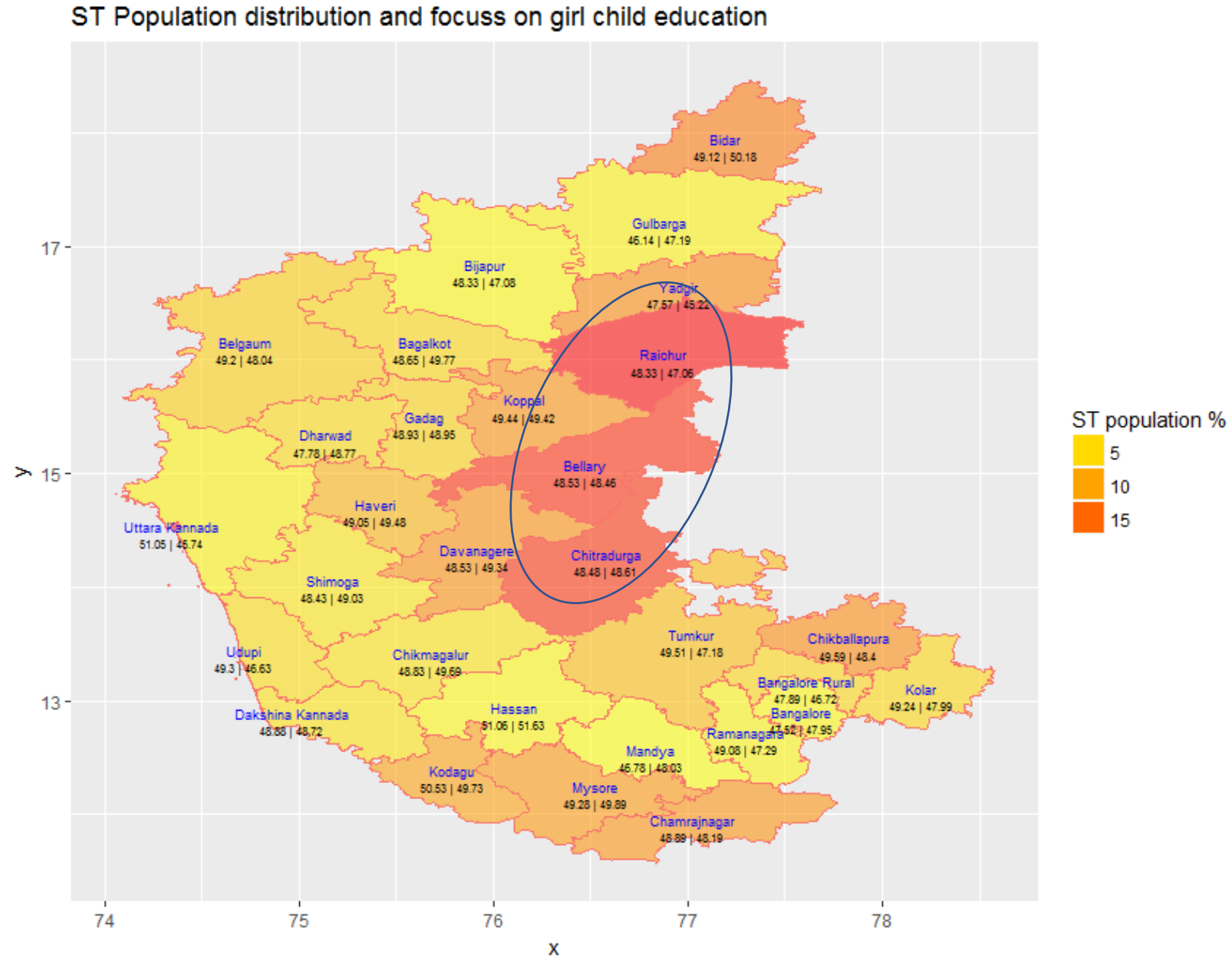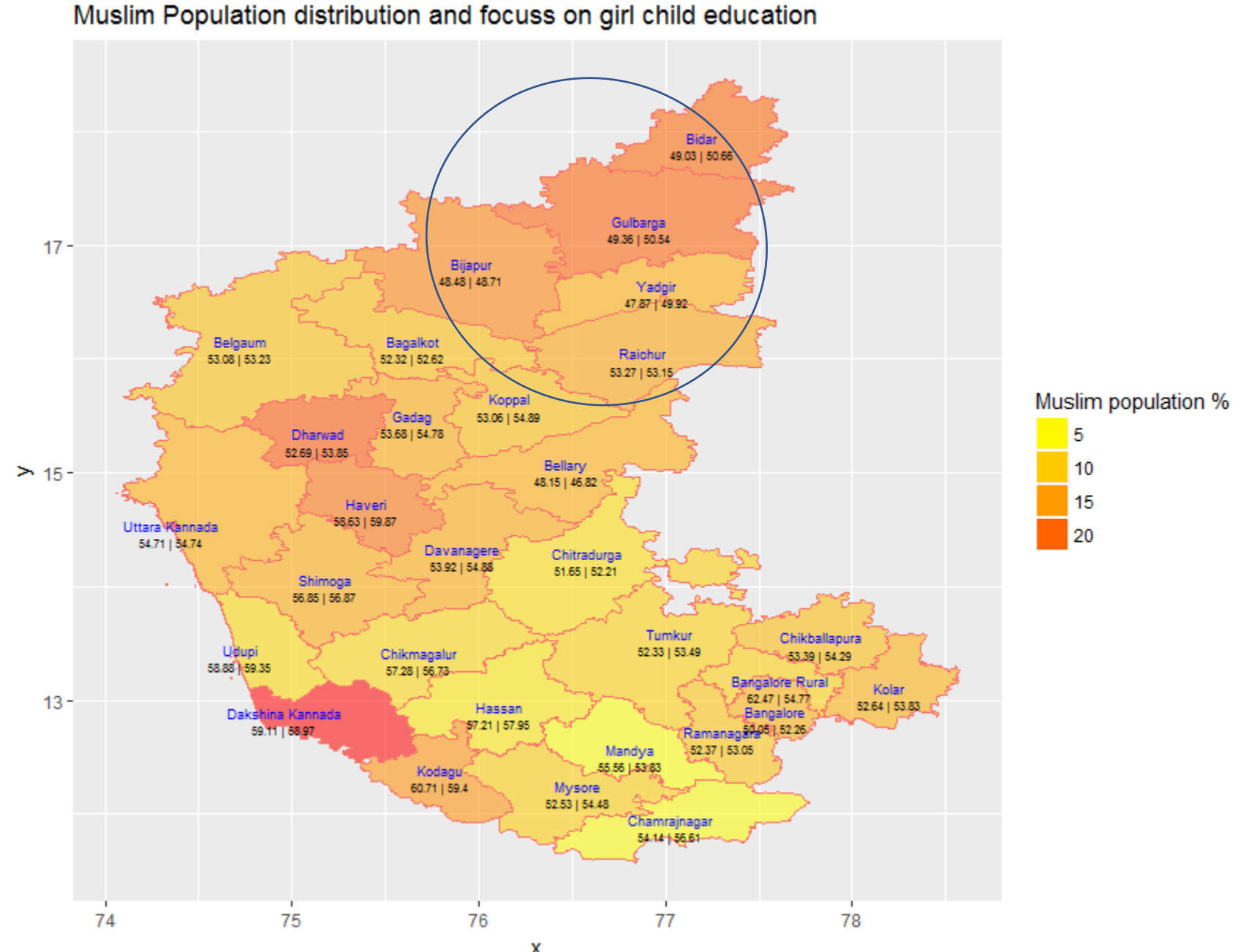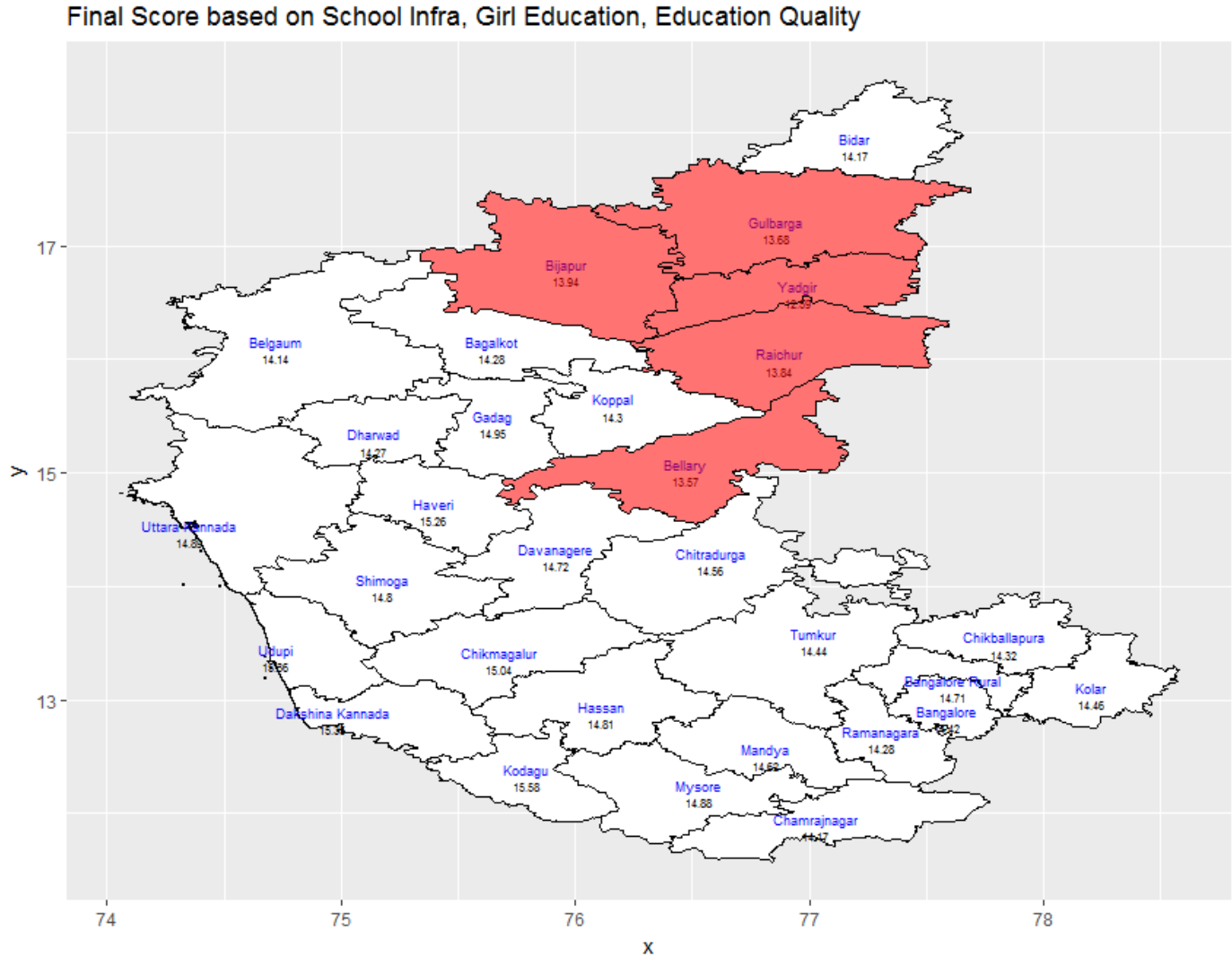
Muslim communities in most of the district have a high proportional of girls been sent to school but this is relatively lower in Bijapur, Bidar, Gulbarga and Yadigr.

Muslim Population distribution and focuss on girl child education

Muslim population %

	5
	10
	15
	20

Bidar 49.03 | 50.66
Gulbarga 49.36 | 50.54
Bijapur 48.48 | 48.71
Yadgir 47.87 | 49.92
Belgaum 53.08 | 53.23
Bagalkot 52.32 | 52.62
Raichur 53.27 | 53.15
Koppal 53.06 | 54.89
Gadag 53.68 | 54.78
Dharwad 52.69 | 53.85
Bellary 48.15 | 46.82
Haveri 58.63 | 59.87
Uttara Kannada 54.71 | 54.74
Davanagere 53.92 | 54.88
Chitradurga 51.65 | 52.21
Shimoga 56.85 | 56.87
Tumkur 52.33 | 53.49
Chikballapura 53.39 | 54.29
Udupi 58.88 | 59.35
Chikmagalur 57.28 | 56.73
Bangalore Rural 62.47 | 54.77
Kolar 52.64 | 53.83
Bangalore 50.05 | 52.26
Dakshina Kannada 59.11 | 58.97
Hassan 57.21 | 57.95
Ramanagara 52.37 | 53.05
Mandya 55.56 | 53.83
Kodagu 60.71 | 59.4
Mysore 52.53 | 54.48
Chamrajnagar 54.14 | 56.61

Overall Scores – Recommended States for Phase 1

Final Score based on School Infra, Girl Education, Education Quality

Sr. No.	District	sch_infr_total_score	girl_edu_total	educ_qua_total_score	total_score
1	Yadgir	7.58	3.74	1.27	12.59
2	Bellary	7.90	3.84	1.83	13.57
3	Gulbarga	7.77	3.87	2.04	13.68
4	Raichur	8.04	3.90	1.91	13.84
5	Bijapur	8.00	3.84	2.09	13.94
6	Belgaum	8.18	4.05	1.92	14.14
7	Bidar	8.03	3.97	2.17	14.17
8	Chamrajnagar	8.25	3.98	1.94	14.17
9	Dharwad	7.90	3.99	2.37	14.27
10	Bagalkot	8.44	3.96	1.87	14.28
11	Ramanagara	8.40	3.92	1.95	14.28
12	Koppal	8.22	3.98	2.10	14.30
13	Chikballapura	8.30	3.98	2.04	14.32
14	Bangalore	8.00	3.91	2.50	14.42
15	Tumkur	8.41	3.89	2.15	14.44
16	Kolar	8.40	3.96	2.10	14.46
17	Chitradurga	8.43	3.93	2.20	14.56
18	Mandya	8.60	3.93	2.09	14.62
19	Bangalore Rural	8.50	4.06	2.14	14.71
20	Davanagere	8.42	4.01	2.29	14.72
21	Shimoga	8.44	4.06	2.30	14.80
22	Hassan	8.49	4.13	2.19	14.81
23	Mysore	8.67	4.02	2.20	14.88
24	Uttara Kannada	8.60	3.94	2.35	14.89
25	Gadag	8.67	4.01	2.28	14.95
26	Chikmagalur	8.67	4.07	2.30	15.04
27	Haveri	8.84	4.10	2.31	15.26
28	Dakshina Kann	8.99	4.11	2.25	15.35
29	Udupi	8.80	4.09	2.47	15.36
30	Kodagu	9.04	4.17	2.37	15.58

Based on the overall score Yadgir, Bellary, Gulbarga, Raichur and Bijapur must be focused in the Phase 1 to improve primary education