

Population de la ville de Saint-Quentin :

```
SELECT * FROM POPULATION
WHERE NOM LIKE 'Saint-Quentin';
```

	⇅ CODEREGION	⇅ CODEDEPT	⇅ NOM	⇅ POP
1	32	2	Saint-Quentin	57482

Villes de plus de 30.000 habitants :

```
SELECT Nom,Pop from population
where pop > 30000 AND CODEDEPT = '59';
```

	⇅ NOM	⇅ POP
1	Cambrai	33609
2	Douai	41941
3	Dunkerque	90721
4	Lille	238381
5	Marcq-en-Baroeul	40246
6	Maubeuge	30722
7	Roubaix	96196
8	Tourcoing	95965
9	Valenciennes	44684
10	Villeneuve-d'Ascq	63732
11	Wattrelos	41580

Statistiques sur la table :

	Nom	Valeur
1	NUM_ROWS	35798
2	BLOCKS	150
3	AVG_ROW_LEN	23
4	SAMPLE_SIZE	35798
5	LAST_ANALYZED	22/02/18
6	LAST_ANALYZED_SINCE	22/02/18

La table compte 35798 lignes, pour 150 blocks alloués. La longueur moyenne d'une ligne est de 23.

Statistiques sur les colonnes :

⇅ TABLE_NAME	⇅ COLUMN_NAME	⇅ NUM_DISTINCT	LOW_VALUE	HIGH_VALUE	⇅ DENSITY	⇅ NUM_NULLS	⇅ NUM_BUCKETS	⇅ LAST_ANALYZED	⇅ SAMPLE_SIZE	⇅ GLOBAL_STATS	⇅ USER_STATS	⇅ AVG_COL_LEN
POPULATION	CODEREGION	13 3131	3934		0, 0769230769230769	0	1	22/02/18	35798	YES	NO	3
POPULATION	CODEDEPT	96 3120	3935		0, 0104166666666667	0	1	22/02/18	35798	YES	NO	3
POPULATION	NOM	33405 41617374	CE6C652D64652D5365696E		0, 0000345346159755229	0	254	22/02/18	35798	YES	NO	13
POPULATION	POP	5815 80	C3302B2F		0, 000769269183444643	0	254	22/02/18	35798	YES	NO	4

On peut observer les valeurs minimum et maximum pour chaque colonne.

Les colonnes nom et pop sont des « Height balanced histogram ».

Les tailles moyennes de colonnes sont de 3 pour CODEREGION, 3 pour CODEDEPT, 13 pour NOM, 4 pour POP.

Procédure CreateHistogram

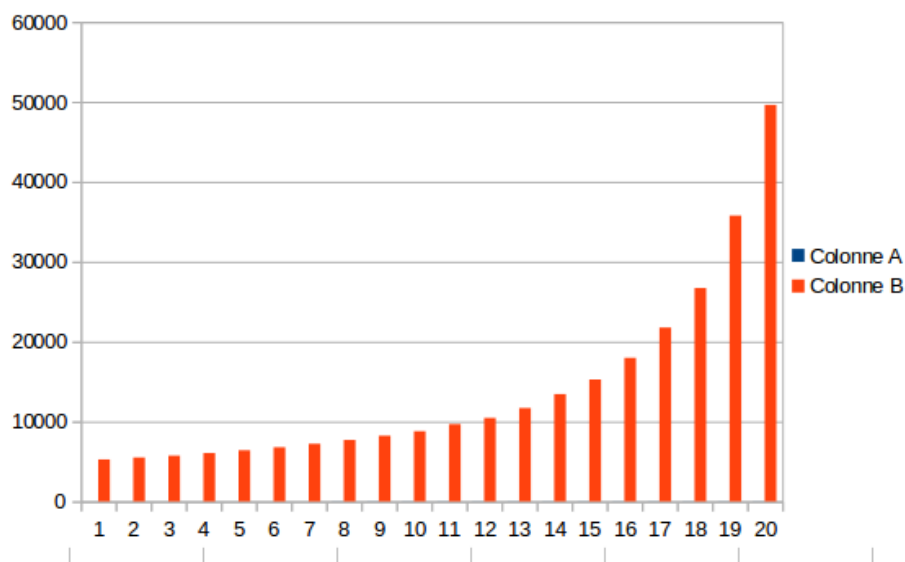
```
CREATE OR REPLACE PROCEDURE CREATEHISTOGRAM ( NOMHISTO IN VARCHAR2,  
NBBUCKETS IN NUMBER, BORNE_INF IN NUMBER, BORNE_SUP IN NUMBER ) AS  
rowCount NUMBER;  
nbLineByBucket NUMERIC;  
bucketIndex NUMBER;  
lastPopValue NUMBER;  
CURSOR pop_cursor IS SELECT ROWNUM,POP FROM POPULATION WHERE POP  
BETWEEN  
BORNE_INF AND BORNE_SUP;  
BEGIN  
IF NBBUCKETS > 0 THEN  
bucketIndex:= 1;  
SELECT COUNT(*) INTO rowCount FROM POPULATION WHERE POP BETWEEN  
BORNE_INF AND BORNE_SUP;  
IF rowCount > 0 THEN  
nbLineByBucket := CEIL(rowCount/NBBUCKETS);  
lastPopValue := -1;  
FOR cursor_raw in pop_cursor  
LOOP  
IF MOD(cursor_raw.ROWNUM,nbLineByBucket) = 0 THEN  
INSERT INTO HISTOPOP VALUES (NOMHISTO,bucketIndex,cursor_raw.POP);  
lastPopValue := -1;  
bucketIndex := bucketIndex+1;  
ELSE  
lastPopValue := cursor_raw.POP;  
END IF;  
END LOOP;  
IF lastPopValue > -1 THEN  
INSERT INTO HISTOPOP VALUES (NOMHISTO,bucketIndex,lastPopValue);  
END IF;  
END IF;  
END IF;  
END CREATEHISTOGRAM;  
/
```

Graphiques et analyses de la question 5

20 buckets, population de 5000 à 50000 :

NOMHISTO	BUCKETINDEX	HIGHVALUE
1 histoPop5000to5000	1	5249
2 histoPop5000to5000	2	5490
3 histoPop5000to5000	3	5729
4 histoPop5000to5000	4	6047
5 histoPop5000to5000	5	6393
6 histoPop5000to5000	6	6772
7 histoPop5000to5000	7	7209
8 histoPop5000to5000	8	7681
9 histoPop5000to5000	9	8220
10 histoPop5000to5000	10	8766
11 histoPop5000to5000	11	9659
12 histoPop5000to5000	12	10439
13 histoPop5000to5000	13	11678
14 histoPop5000to5000	14	13423
15 histoPop5000to5000	15	15245
16 histoPop5000to5000	16	17959
17 histoPop5000to5000	17	21743
18 histoPop5000to5000	18	26695
19 histoPop5000to5000	19	35761
20 histoPop5000to5000	20	49600

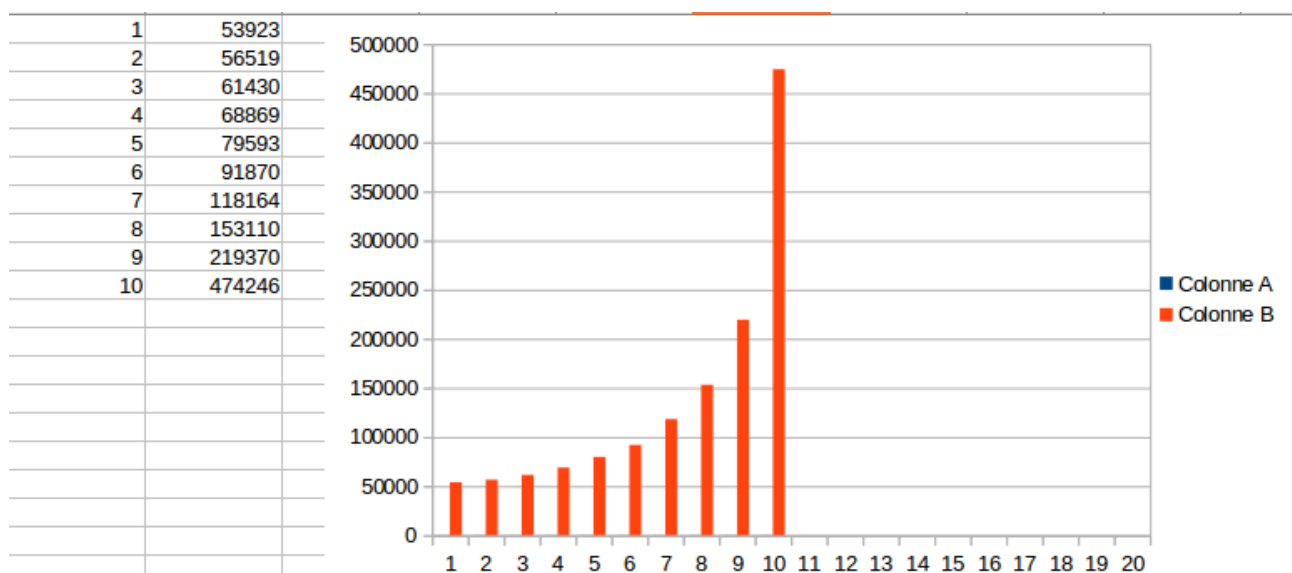
1	5249
2	5490
3	5729
4	6047
5	6393
6	6772
7	7209
8	7681
9	8220
10	8766
11	9659
12	10439
13	11678
14	13423
15	15245
16	17959
17	21743
18	26695
19	35761
20	49600



On remarque, avec 20 buckets, que la majorité des villes sont des villes de taille petite (entre 5000 et 20000 habitants, représentant 17 buckets) comparativement aux «moyennes» villes (entre 20000 et 50000 habitants pour 3 buckets).

10 buckets, population de 50000 à 500000 :

21	histoPop50000to...	1	53923
22	histoPop50000to...	2	56519
23	histoPop50000to...	3	61430
24	histoPop50000to...	4	68869
25	histoPop50000to...	5	79593
26	histoPop50000to...	6	91870
27	histoPop50000to...	7	118164
28	histoPop50000to...	8	153110
29	histoPop50000to...	9	219370
30	histoPop50000to...	10	474246



On remarque, avec 10 buckets, que pour un échantillon de villes entre 50 000 et 500 000 habitants, la plupart des échantillons se trouvent dans les villes moyennes.