

## **Modern Classification: Principles and Documentation - POPULATION**

**Authors: Gijs Kessler and Andrei Markevich**

### **Introduction**

Historical statistics are coined in the categories and terms of the period – what we refer to within the framework of the Electronic Repository of Russian Historical Statistics as ‘historical classifications’. These categories can even differ within one and the same period, due to the existence of different spelling variants and synonyms. Particularly for the earlier benchmark-years 1795 and 1858 this is often the case. Standardising these categories across data-sets in order to maximise the comparability of the data is one of the central aims of the Electronic Repository of Russian Historical Statistics (ERRHS).

To achieve this, we have:

- (a) unified and systematised the 'historical' categories used in the sources and;
- (b) matched and marked them accordingly to categories used in accepted modern, and highly standardised systems of classification (NACE, CLEMS, PST and others).

in which:

- (a) The standardisation of historical categories into what we will refer to as ‘unified historical classifications’ serves the purpose of making data comparable within one and the same period and facilitates querying the data.

- (b) the reliance on modern classifications to tag the data resolves the issue of comparability over time. Unifying 'historical classifications' does not render data comparable between cross-sections. Different grouping principles might have been used, and the meaning of terms and categories apparently similar, can in fact have changed over the years. Coding the data for different years using one and the same classification scheme does, however, solve this problem.

On the Ristat-portal users can query the database using the historical or the modern classification, and aggregate the data at different levels of hierarchy. This query tool is accessed from the left-hand menu at ‘STATISTICS/TOPICS’. A radio-button allows the user to toggle between historical and modern classification, changing the set of categories available for selection under ‘4. Choose indicators’.

Full disaggregated data-sets in database format (one row=one record) with simultaneous historical and modern classification can be downloaded from the left-hand menu at ‘STATISTICS/FILE CATALOGUE’. The historical classification is in the fields ‘HistClass1’, ‘HistClass2’.....’HistClass10’. The modern classification is in the fields ‘Class1’, ‘Class2’.....’Class10’. Below, we expand on the methodology used in applying modern classifications.

### **General principles of the application of modern classifications**

Modern classifications are, as a rule, hierarchical, and consist of nested categories, which allow for the classification of each and every historical variable, even if at different levels of detail. Branches and sub-branches of industry can be taken as a good example. In case a certain figure cannot be matched to a single sub-branch of industrial production, it might well be feasible to determine the more general branch of production that it relates to, one level up in the hierarchy.

However, such a procedure rarely allows one to reconstruct all categories of the modern classification, because the necessary level of detail of the data to do so might simply not be available in the source. Only if each historical category matches to only one category in the modern classification (a relationship of many-to-one), and all necessary historical categories are identifiable and available in the source ('functional completeness') can we speak of a full reconstruction of a category from the modern classification. This is usually the case only at the higher, as well as at the lower levels of aggregation, whereas at intermediate levels the mismatches are a more vexing problem.

This notwithstanding, we have categorised all data in the database in terms of a modern classification, either through full reconstructions, or by distributing the available historical categories among modern 'baskets'. In the next sections, we will specifically elaborate on our methodology with regards to the modern classification of ERRHS' data on population.

### **Population - 'simplified historical classification'**

In choosing the modern classifications applied in the project we have consistently set the aim of using the system of classification most widely accepted in the corresponding field of knowledge. No generally accepted modern classification for data on population appears to exist, however, likely because of the self-explanatory nature of many of the categories involved. Nevertheless, in order to be able to line up data over time and compare across benchmarks, we need a unified system of categories, or groupings of categories. To this aim we have created a 'simplified historical classification', a harmonization of categories over time using generally accepted common parlance terms.

The categories used in this system of classification are specified in the paragraphs below for each subtopic of POPULATION (the underlined words are the actual terms used in the database):

#### **1.01 - Population by sex**

Male; Female; Both sexes; Unknown

#### **1.02 - Population by age**

The following larger age-groups are distinguished (Class1):

Up to six years

7 to 10 years

11 years

12 years

13 to 14 years

15 to 16 years

17 to 54 years

55 to 59 years

60 years

61 to 64 years

65 years

66 to 74 years

75 years and over

Unknown age

If the data allow for this, a further subdivision of these groups is available at the next hierarchical level (Class2):

<u>0</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>
<u>7</u>	<u>8</u>	<u>9</u>	<u>10</u>	<u>11</u>	<u>12</u>	<u>13</u>
<u>14</u>	<u>15</u>	<u>16</u>	<u>17</u>	<u>18</u>	<u>19</u>	<u>20</u>
<u>21</u>	<u>22</u>	<u>23</u>	<u>24</u>	<u>25</u>	<u>26</u>	<u>27</u>
<u>28</u>	<u>29</u>	<u>30</u>	<u>31</u>	<u>32</u>	<u>33</u>	<u>34</u>
<u>35</u>	<u>36</u>	<u>37</u>	<u>38</u>	<u>39</u>	<u>40</u>	<u>41</u>
<u>42</u>	<u>43</u>	<u>44</u>	<u>45</u>	<u>46</u>	<u>47</u>	<u>48</u>
<u>49</u>	<u>50</u>	<u>51</u>	<u>52</u>	<u>53</u>	<u>54</u>	<u>55</u>
<u>56</u>	<u>57</u>	<u>58</u>	<u>59</u>	<u>60</u>	<u>61</u>	<u>62</u>
<u>63</u>	<u>64</u>	<u>65</u>	<u>66</u>	<u>67</u>	<u>68</u>	<u>69</u>
<u>70</u>	<u>71</u>	<u>72</u>	<u>73</u>	<u>74</u>	<u>75</u>	<u>76</u>
<u>77</u>	<u>78</u>	<u>79</u>	<u>80</u>	<u>81</u>	<u>82</u>	<u>83</u>
<u>84</u>	<u>85</u>	<u>86</u>	<u>87</u>	<u>88</u>	<u>89</u>	<u>90</u>
<u>91</u>	<u>92</u>	<u>93</u>	<u>94</u>	<u>95</u>	<u>96</u>	<u>97</u>
<u>98</u>	<u>99</u>	<u>100&gt; years</u>	<u>Unknown age</u>			

Where possible, finally, the data are subdivided by sex (Class3):

Male; Female; Sex not specified; Unknown and by Rural; Urban; Rural/urban status unspecified; No rural-urban breakdown (Class4).

**1.03 - Population urban/rural**

Our modern classification of population data on residence rests on two pillars. On the one hand we have accepted the division between rural and urban as given in the source. The reason for this is entirely pragmatic - imposing our own classification of urban and rural would require data on the number of inhabitants of all population centres, data which are mostly not available. The consequence if, of course, that we cannot correct for slightly diverging definitions over time of what was rural and urban. In Tsarist Russia, for example, urban status was exclusively a matter of legal status, without taking into account the social and economic characteristics of the place. In Soviet times, urban status reflected a certain population size and an occupational structure which relied predominantly on non-agrarian pursuits.

For those years and regions where we have disaggregated data for individual towns we have additionally calculated the subdivision of the urban population into those resident in towns with under 5000 inhabitants and in towns with over 5000 inhabitants.

Where available, we provide separate data for Male and Female population.

This results in the following classification grid:

Class 1	<u>Urban</u> ; <u>Rural</u> ; <u>Rural/urban status unspecified</u> ; <u>No rural-urban breakdown</u>
Class 2	<u>&lt;5000 inhabitants</u> ; <u>≥5000 inhabitants</u>
Class 3	<u>Male</u> ; <u>Female</u> ; <u>Sex not specified</u> ; <u>Unknown</u>

**1.04 - Population by confession**

Data on the religious background of the population we have standardised using a simplified classification of religious denominations (Class1). Where possible, we have also differentiated by sex (Class2) between Male; Female; Both sexes; Unknown.

The correspondence between historical and modern categories for Class1 is given in the table below:

<i>Modern (Class1)</i>	<i>1795 (HistClass1)</i>	<i>1858 (HistClass1)</i>	<i>1897 (HistClass1)</i>	<i>1959</i>	<i>2002</i>
<u>Protestants</u>	Lutheran, Reformed, Catholic; Protestant	Lutheran and Reformed; Mennonite; Protestant; Anglican	Anglican; Lutheran; Mennonite; Reformed; Baptist	NO DATA	
<u>Catholics</u>	Roman Catholic Armenian	Roman Catholic Armenians; Roman Catholic	Armenian Catholic; Roman Catholic		
<u>Orthodox</u>	Orthodox	Orthodox; Common Faith	Orthodox; Orthodox and Common Faith		
<u>Old Believers</u>	Schismatic;	Schismatic/ Priestless sects that do not recognize marriage and do not pray for Tsar; Schismatic / Priestless sects that recognize marriage and pray for Tsar; Schismatic / Molokan, Doukhobor, Ikonobortsy; Schismatic/ Accepting priesthood; Schismatic/ Skoptsy, Khlysty and other adherents of Skopchestvo; Schismatic/ Sabbatarian or Judaizer	Old Believers and those deviating from Orthodoxy		
<u>Other Christians</u>	Armenian	Armenian-Gregorian	Armenian Gregorian; Other Christian faiths; Followers of other Christian faiths		
<u>Buddhists</u>	-	Lamaism; Buddhism	Buddhist and Lamaist		
<u>Jews</u>	-	Judaism	Jewish		
<u>Muslims</u>	Bukharan; Muslim; Murzas;	Muslim	Muslim		
<u>Others</u>	Fetishist; Pagan	Fetishism; Karaism; Shamanism	Karaite; Other non-Christian faiths		
<u>Unknown</u>	Non-Christian; Steppe nomads	All faiths			

### 1.05 - Population by estate/social group

No modern classification is applied for this subtopic, because the historical categories concerned (estate for 1795, 1858 and 1897; soviet official classes for 1959; number of pensioners for 2002) cannot sensibly be compared over time using a single set of criteria. Attempts have been made to come up with a classification scheme for such purposes, like HISCLASS and HISCAM, but they all relate to data on professions:

HISCLASS: <https://collab.iisg.nl/web/hisco/hisclass>

HISCAM: <https://collab.iisg.nl/web/hisco/hiscam>

### 1.06 - Number of births

The main variable for this subtopic - the number of births is classified in field Class1 as Number of Births. Additionally, where possible it is distinguished in Class2 between Male; Female; Both sexes; Unknown, and between Urban; Rural; Unknown; All in Class3, with a further subdivision for the urban population in Class4 between <5000 inhabitants; ≥5000 inhabitants.

Class 1	<u>Number of births</u>
Class 2	<u>Male</u> ; <u>Female</u> ; <u>Sex not specified</u> ; <u>Unknown</u>
Class 3	<u>Urban</u> ; <u>Rural</u> ; <u>Rural/urban status unspecified</u> ; <u>No rural-urban breakdown</u>
Class 4	<u>&lt;5000 inhabitants</u> ; <u>≥5000 inhabitants</u>

Further indicators for this subtopic, such as the number of twins, multiple births, and illegitimate births, are classified according to the following scheme:

Class1 (RUS)	Class1 (EN)
<u>близнецы и многоплодная беременность</u>	<u>Twins and multiple births</u>
<u>внебрачные дети</u>	<u>Illegitimacy</u>
<u>пороки развития</u>	<u>Congenital disorders</u>
<u>подкидыши</u>	<u>Abandoned infants</u>
Class2 (RUS)	Class2 (EN)
<u>мужчины</u> ; <u>женщины</u> ; <u>без разделения по полу</u> ; <u>неизвестно</u>	<u>Male</u> ; <u>Female</u> ; <u>Sex not specified</u> ; <u>Unknown</u>

### 1.07 - Number of deaths

The main variable for this subtopic - the number of deaths is classified in field Class1 as Number of deaths. Additionally, where possible it is distinguished in Class2 between Male; Female; Sex not specified; Unknown, and between Urban; Rural; Unknown; No rural-urban breakdown in Class3, with a further subdivision for the urban population in Class4 between <5000 inhabitants; ≥5000 inhabitants.

Class1	<u>Number of deaths</u>
Class2	<u>Male</u> ; <u>Female</u> ; <u>Sex not specified</u> ; <u>Unknown</u>
Class3	<u>Urban</u> ; <u>Rural</u> ; <u>Rural/urban status unspecified</u> ; <u>No rural-urban breakdown</u>
Class4	<u>&lt;5000 inhabitants</u> ; <u>≥5000 inhabitants</u>

Data on the age at death are classified according to the following scheme:

Class1	<u>Deaths by age</u>
Class2	<u>Male</u> ; <u>Female</u> ; <u>Sex not specified</u> ; <u>Unknown</u>
Class3	<u>Urban</u> ; <u>Rural</u> ; <u>Rural/urban status unspecified</u> ; <u>No rural-urban breakdown</u>
Class4	<u>Up to 14 years</u> ; <u>15-59 years</u> ; <u>60 years and over</u> ; <u>Unknown age</u>

Further indicators for this subtopic, such as the number of sudden deaths from accidents, missing persons etc. are classified according to the following scheme:

Class1	<u>Sudden deaths and accidents</u> ; <u>Missing persons</u> ; <u>Deaths from epidemic diseases</u> ; <u>Stillbirths</u>
Class2	<u>Male</u> ; <u>Female</u> ; <u>Sex not specified</u> ; <u>Unknown</u>
Class3	<u>Urban</u> ; <u>Rural</u> ; <u>Rural/urban status unspecified</u> ; <u>No rural-urban breakdown</u>
Class4	<u>&lt;5000 inhabitants</u> ; <u>≥5000 inhabitants</u>

**1.08 - Number of marriages**

Class 1	<u>Number of marriages</u>
Class 2	<u>Urban; Rural; Rural/urban status unspecified; No rural-urban breakdown</u>
Class 3	<u>&lt;5000 inhabitants; ≥5000 inhabitants</u>

**1.09 - Marital status**

Data on marital status are classified according to the following scheme, distinguishing between Married; Single; Unknown marital state in Class1, with a further subdivision in Class2 of the single population between Not previously married; Widowed; Divorced. Where possible, we also distinguish by sex (Class3).

Class 1	<u>Married; Single; Unknown marital state</u>
Class 2	<u>Not previously married; Widowed; Divorced</u>
Class 3	<u>Male; Female; Sex not specified; Unknown</u>

### 1.10 - Education

For classifying the data on the educational attainment of the population we rely on a simplified historical classification, which distinguishes between four broad categories of schooling (Class1) - Illiterate; Primary and secondary school; Higher education; Level of education unspecified. This simplified classification does not take account of differences in the definition of these three categories over time.

The correspondence between historical and modern categories for Class1 is given in the table below:

Modern Classification (Class1)	1795	1858	1897	1959	2002
<u>Illiterate</u>	NO DATA		Illiterate	Illiterate	Illiterate
<u>Primary and secondary school</u>	NO DATA		Literate, not studying at higher or secondary educational institutions; Literate, studying at secondary educational institutions; Literate, studying at specialized secondary educational institutions	Primary and secondary	Professional primary; Professional secondary; General primary; General secondary; General basic
<u>Higher education</u>	NO DATA		Literate, studying at universities and other higher educational institutions; Literate, studying at higher specialized and technical colleges; Literate, studying at higher military academies	Incomplete post secondary; Post secondary	Professional post-graduate education; Professional post-secondary education; Professional incomplete post-secondary
<u>Level of education unspecified</u>	NO DATA				Level of education unspecified; Literate, without primary and secondary education

Where possible, we also distinguish by sex between Male; Female; Sex not specified; Unknown (Class2), and between Urban; Rural; Rural/urban status unspecified; No rural-urban breakdown (Class3).

\*\*\*

October 2020