ELSEVIER
CrossMark
Available online at www.sciencedirect.com
ScienceDirect
Procedia Engineering 162 (2016) 585 – 592
Procedia Engineering
www.elsevier.com/locate/procedia

International Conference on Efficient & Sustainable Water Systems Management toward Worth Living Development, 2nd EWaS 2016

# Evaluating *Google Trends* as a Tool for Integrating the '*Smart Health*' Concept in the Smart Cities' Governance in USA

Alexia Sampri[a], Amaryllis Mavragani[a], Konstantinos P. Tsagarakis[a,*]

[a]*Business and Environmental Technology Economics Lab, Department of Environmental Engineering, School of Engineering, Democritus University of Thrace, Xanthi 67100, Greece*

**Abstract**

The aim of this paper is to introduce the methodology of using online search traffic data in order to integrate the public's online behavior in Smart Health; a concept that is currently rising concerning the health factor of Smart Cities. We use normalized data from Google Trends from January 2013 to December 2015 in the US, aiming at exploring the change in interest in various medical terms, and examine if Google Trends is a possible tool for evaluating health search queries by nowcasting the public's online interest. The results show that Google Trends' data can be used for measuring the public's interest in health related terms, in order to assist with the evaluation of 'Smart Health'.


Peer-review under responsibility of the organizing committee of the EWaS2 International Conference on Efficient & Sustainable Water Systems Management toward Worth Living Development

*Keywords:* Big Data; Google Trends; Online Behavior; Smart Cities; Smart Health

## 1. Introduction

As Smart Cities are becoming all the more popular in science and governance [1] over the last decade, large amounts of data are needed in order to access and evaluate the six pillars of a Smart City, namely 'Smart Economy', 'Smart People', 'Smart Environment', 'Smart Mobility', 'Smart Living', and 'Smart Governance' [2]. Smart Cities are monitored by different kinds of sensors for their evaluation [3], thus allowing the constant gathering of large amount of data [3-4].

* Corresponding author. Tel.: +30 2541079397; fax: +30 2541079397.
*E-mail address:* ktsagar@env.duth.gr

1877-7058 
Peer-review under responsibility of the organizing committee of the EWaS2 International Conference on Efficient & Sustainable Water Systems Management toward Worth Living Development
doi:10.1016/j.proeng.2016.11.104

A rising concept is that of Smart Health [5-6], aiming at using mobile health data in Smart Cities [7]. In order for this concept to be integrated in the Smart Cities' performance, large amount of data, namely Big Data, are needed [5], that could assist with the government's assessment of the health care system's issues [7] and in the evaluation of Smart Cities in general and public health in specific [8]. For example, in the US, where these vast amounts of information are analyzed to assist in *"clinical analytics"* [9].

Big Data have been used in research in the past for the evaluation of the public interest in health issues, as they have been suggested to be valuable in the subject of medicine, and helpful in analyzing patients' data [10] and in health matters in several topics [11]. Google Trends is a tool that is all the more integrated in scientific research in general and health related issues in specific [11]. To name a few, it has been used in order to assess epilepsy related searches [12], sexually transmitted diseases [13], Ebola related search queries during the 2014 outbreak [14], and to relate pertussis searches and incidence [15].

Big Data in general can be of value for the governments and policy makers with the analysis of online search queries. The online public interest could be assessed centrally and then applied by region in order to improve health locally i.e. integrate the analysis in Smart Cities. The aim of this paper is to examine how Google Trends' data can be valuable in assisting with the evaluation of the interest in health issues in the US. We choose the terms 'Asthma', 'Lyme disease', 'Melanoma', 'COPD', and 'Salmonella', as representative terms of the public's general interest. The rest of the paper is structured as follows: Section 2 consists of the research methodology, in section 3 the results are presented and discussed, and section 4 consists of the overall conclusions.

## 2. Methodology

We use the Google Trends' [16] hits' data from January 2013 to December 2015 to analyze the change in the online interest in the terms are 'Asthma', 'Lyme disease', 'Melanoma', 'COPD', and 'Salmonella' in the US. Data are normalized over each selected period and are downloaded online in '*.csv' format.

Furthermore, we analyze each term's interest by State and we proceed to categorize the interest in terms of normalized hits in 5 groups: very high interest (80-100), high interest (60-80), moderate interest (40-60), fair interest (20-40), and poor interest (0-20). In addition, we provide the visualization of the data and examine if any consistencies exist amongst the States' rankings of online interest and reported incidents.

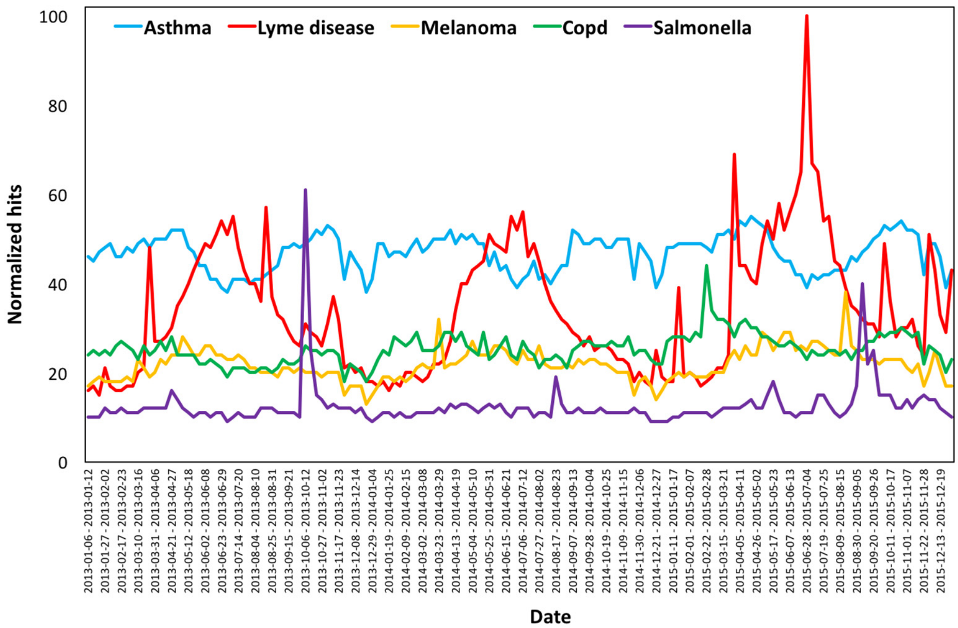


Fig. 1. Normalized hits in 'Asthma', 'Lyme disease', 'Melanoma', 'COPD' and 'Salmonella' in the US from 2013 to 2015.

## 3. Results and Discussion

Figure 1 shows the normalized hits in Google in the terms ‘Asthma’, ‘Lyme disease’, ‘Melanoma’, ‘COPD’, and ‘Salmonella’ in the US from January 2013 to December 2015.

Following, we present the visualizations of the normalized searches by State. Figure 2 consists of the Google searches in the 50 States in (a) Asthma, (b) Lyme disease, (c) Melanoma, (d) COPD, and (e) Salmonella from January 2013 to December 2015. We divide the interest in terms of normalized hits in 5 groups: very high interest (80-100), high interest (60-80), moderate interest (40-60), fair interest (20-40), and poor interest (0-20), and we proceed to analyze each term’s interest by State, in addition to the cross-reference of the States’ highest and lowest interest in the five examined terms.

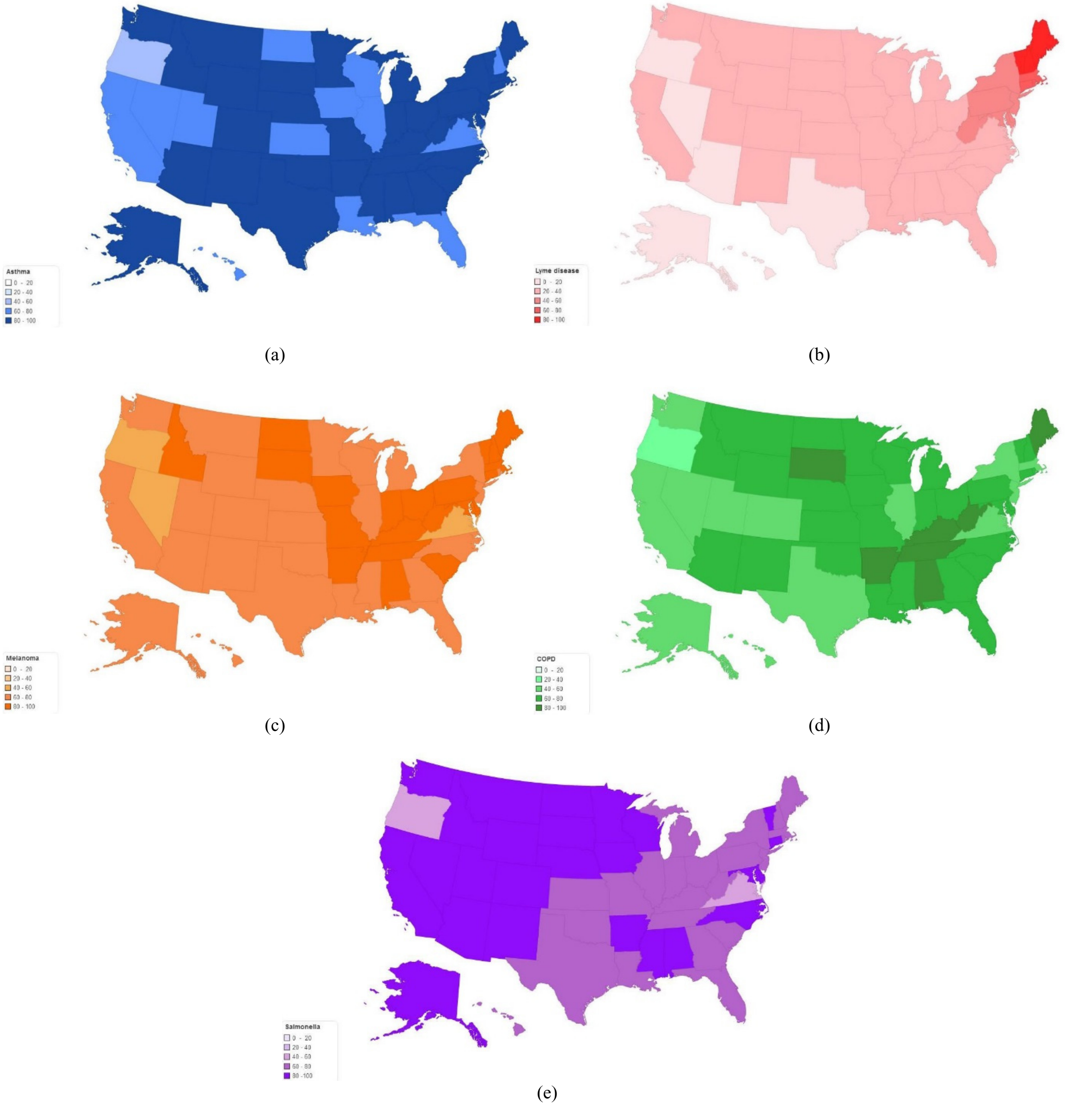


Fig. 2. Online interest by State in (a) Asthma, (b) Lyme disease, (c) Melanoma, (d) COPD, and (e) Salmonella from 2013 to 2015.

In Google Trends, when search volumes are not high enough, the State's scoring is '0'. What is interesting in this study is that Google Trends provides data for all States in all five examined diseases, thus we observe that the interest is high enough to be evaluated in all States throughout the examined period. In order to further elaborate on the usefulness of Google Trends' data, we compare the searches with the geographical patterns and the reported cases of the examined diseases.

For 'Lyme disease', as reported by the US' Center for Disease Control (CDC) and prevention [17], we observe that there exist obvious similarities in the patterns of Google searches and incidence of the disease, mainly concentrated in the Northeast and upper Midwest [18].

A similar pattern can be observed for the term 'COPD'. When compared to the (most recent) map provided by CDC in 2011 [19], we observe that the Southeastern States are the ones with the most incidents of COPD, in line with the searches for this term in Google. Similar results can be derived for Google searches in the term 'Melanoma' and the death rates resulting from this disease in 2013, though not for the incidence rates [20].

It is not the case, though, that it is at all times easy to raise conclusions from comparing the maps of incidence with Google hits, as, for example, for the term 'Asthma', since it is mostly reported in big cities [21]. For example, the 5 cities with the highest reported incidents of Asthma in 2015 were Memphis (Tennessee), Richmond (Virginia), Philadelphia (Pennsylvania), Detroit (Michigan), and Oklahoma City (Oklahoma) [21]. The aforementioned States rank 5$^{th}$, 49$^{th}$, 11$^{th}$, 26$^{th}$, and 25$^{th}$ in Google searches, respectively.

Furthermore, Google Trends' hits cannot always be in line with the incidents when the medical term is not univocally defined. For example, in the case of Salmonella, there exist many different serotypes (more than 2500), only less than 100 of which are for concern for human health [22], thus Google searches for this disease should not be directly linked to incidents of Salmonella [23] in humans, and further analysis requires caution.

The overall online interest in the five terms over the examined period in the US, based on the mean of the percentized monthly hits' averages (%) is as follows: Asthma (33.84%), Lyme disease (23.53%), COPD (18.22%), Melanoma (15.56%), and Salmonella (8.85%), as shown in Table 1. It is observed that there is no significant variation during the examined years. The monthly averages (%) of the percentized hits in the five examined terms from 2013 to 2015 can be found in Appendix A.

Table 1. Mean of the monthly averages for 2013, 2014, and 2015, and the total average of percentized online interest.

| Year | Asthma | Lyme disease | Melanoma | COPD | Salmonella |
|---|---|---|---|---|---|
| 2013 | 34.68% | 23.22% | 15.41% | 17.33% | 9.36% |
| 2014 | 34.89% | 21.93% | 15.99% | 18.81% | 8.37% |
| 2015 | 31.96% | 25.45% | 15.26% | 18.50% | 8.81% |
| Total | 33.84% | 23.53% | 15.56% | 18.22% | 8.85% |

The next step in the evaluation of the interest in the five diseases, is to examine whether or not there exist any constistencies in the States' rankings. Tables 2 and 3 consist of the Top Ten and Bottom Ten States in terms of interest in each of the examined keywords. The full table consisting of all the States' rankings in interest in 'Asthma', 'Lyme disease', 'Melanoma', 'COPD', and 'Salmonella' can be found in descending order in Appendix B.

Table 2. Top ten States in hits in 'Asthma', 'Lyme disease', 'Melanoma', 'COPD' and 'Salmonella' in the US.

| Asthma | Lyme disease | Melanoma | COPD | Salmonella |
|---|---|---|---|---|
| Maine | Maine | West Virginia | West Virginia | Idaho |
| Delaware | New Hampshire | Tennessee | Kentucky | South Dakota |
| Kentucky | Vermont | Pennsylvania | Arkansas | North Dakota |
| Connecticut | Rhode Island | Vermont | Tennessee | Montana |
| Tennessee | Connecticut | Maine | Maine | Utah |
| West Virginia | Massachusetts | North Dakota | South Dakota | Wyoming |
| North Carolina | Pennsylvania | Alabama | Alabama | Arkansas |
| Mississippi | Maryland | Delaware | Wyoming | New Mexico |
| Rhode Island | Delaware | New Hampshire | Mississippi | Alaska |
| Maryland | New York | Massachusetts | Indiana | Vermont |

Though there are some consistencies in the Top Ten lists in online interest, e.g. Maine in 4/5 and Delaware & Tennessee in 3/5, we observe that there is no State in all of the lists, meaning that the five terms' Top Ten States lists' cross-section provides no result.

On the other hand, it is significant that we compare Google search data with real incidents. For example, the top 14 States in which 96% of the reported cases in Lyme disease were reported in 2014 were (in alphabetical order) Connecticut, Delaware, Maine, Maryland, Massachusetts, Minnesota, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, Virginia, and Wisconsin [24], that in Google searches rank $1^{st}$ to $11^{th}$, $13^{th}$, $14^{th}$ and $16^{th}$, as shown in Table B1. This indeed indicates that the public's interest is depicted in online search queries.

Table 3. Bottom ten States in hits in 'Asthma', 'Lyme disease', 'Melanoma', 'COPD', and 'Salmonella' in the US.

| Asthma | Lyme disease | Melanoma | COPD | Salmonella |
|---|---|---|---|---|
| Utah | Washington | Illinois | Colorado | Kansas |
| Wisconsin | Colorado | Washington | Texas | Texas |
| Florida | New Mexico | New York | New Jersey | Florida |
| Nevada | Utah | Hawaii | Washington | Tennessee |
| California | Alaska | California | New York | Illinois |
| Iowa | Arizona | Alaska | Hawaii | Michigan |
| Hawaii | Nevada | Texas | Utah | New Jersey |
| Louisiana | Texas | Nevada | California | New York |
| Virginia | Hawaii | Virginia | Virginia | Virginia |
| Oregon | Oregon | Oregon | Oregon | Oregon |

In the Bottom Ten States' lists, though there are examples of States being in more than one list, e.g. Nevada in 3/5, Virginia in 4/5, and Texas in 4/5, the cross-section between the States in the five diseases provides only one result; the State of Oregon which is at the bottom of each disease's list in terms of interest.

In the $2^{nd}$, $3^{rd}$, and $4^{th}$ groups of rankings (see Appendix B), we observe that, again, there is no State that is in the same group of interest ranking in all five diseases.

Based on the above and given the consistencies observed when cross-referencing the terms' searches in Google and the incidents reported in the States, we suggest that Google Trends can be a valuable and reliable tool in nowcasting some diseases' spreading and incidence. The latter could assist with the individual States' monitoring of Google searches in order to predict any possible outbreaks, and be prepared and better handle incidents of diseases.

## 4. Conclusions

Big Data have been used in the past in order to monitor and help with the evaluation of Smart Cities in general and public health in specific. As 'Smart Health', a concept that is currently rising concerning the health factor of Smart Cities, is being integrated in the Smart Cities governance, innovative tools of accessing data and analyzing the public's interest and behavior are needed.

Big Data have already been suggested in the past to be valuable towards this direction. Google Trends, a popular tool to access these kinds of data, has been proven effective in the past in predicting and analyzing several medical and health-related terms' change in online interest. The aim of this paper is to merely provide an example showing that Google Trends could be successfully used in integrating online search traffic data for the developing of Smart Health.

Based on the above, we suggest that Google Trends is indeed a reliable tool in measuring the public's interest in health issues, and could be used in assisting with the integration of the Smart Health concept in the Smart Cities' governance. As each State shows different scorings in the keywords' normalized hits, Google Trends can take into account the diversity in interest in each State, which could assist with the predicting and nowcasting of possible incidents and outbreaks. Google Trends' data on health related terms can be used centrally by analysts and policy makers in order to locally assist with the Smart Cities' organization, governance, and monitoring in the US; an idea that could be also applied to other countries and regions.

## Appendix A. Monthly Averages

Table A1 consists of the monthly averages (%) of the percentized hits in Asthma, Lyme disease, Melanoma, COPD, and Salmonella in the US from 2013 to 2015.

Table A1. Monthly averages (%) of the percentized hits in the US from 2013 to 2015.

| | Asthma | Lyme disease | Melanoma | COPD | Salmonella |
|---|---|---|---|---|---|
| Jan-13 | 39.75% | 15.48% | 15.48% | 20.50% | 8.79% |
| Feb-13 | 39.75% | 13.52% | 15.37% | 21.93% | 9.43% |
| Mar-13 | 37.19% | 19.85% | 15.56% | 18.67% | 8.74% |
| Apr-13 | 35.68% | 21.10% | 16.12% | 17.84% | 9.26% |
| May-13 | 31.80% | 28.01% | 16.97% | 15.65% | 7.58% |
| Jun-13 | 27.47% | 34.40% | 16.67% | 14.40% | 7.07% |
| Jul-13 | 28.65% | 33.16% | 15.97% | 14.76% | 7.47% |
| Aug-13 | 30.80% | 31.16% | 14.86% | 15.04% | 8.15% |
| Sep-13 | 36.23% | 22.60% | 16.17% | 16.92% | 8.08% |
| Oct-13 | 33.07% | 18.78% | 13.00% | 16.53% | 18.62% |
| Nov-13 | 36.56% | 22.49% | 14.26% | 17.37% | 9.32% |
| Dec-13 | 39.18% | 18.07% | 14.49% | 18.43% | 9.84% |
| Jan-14 | 39.79% | 14.43% | 15.67% | 21.44% | 8.66% |
| Feb-14 | 37.94% | 15.81% | 15.81% | 21.94% | 8.50% |
| Mar-14 | 37.56% | 15.99% | 17.95% | 20.06% | 8.45% |
| Apr-14 | 34.39% | 23.26% | 15.28% | 18.44% | 8.64% |
| May-14 | 30.65% | 29.70% | 15.96% | 15.96% | 7.74% |
| Jun-14 | 28.12% | 32.19% | 16.03% | 16.03% | 7.63% |
| Jul-14 | 28.34% | 32.08% | 15.96% | 16.12% | 7.49% |
| Aug-14 | 31.64% | 25.83% | 15.97% | 16.98% | 9.58% |
| Sep-14 | 36.12% | 20.46% | 16.55% | 18.51% | 8.36% |
| Oct-14 | 36.68% | 19.16% | 16.42% | 19.53% | 8.21% |
| Nov-14 | 38.84% | 16.98% | 15.25% | 20.44% | 8.49% |
| Dec-14 | 38.65% | 17.25% | 15.07% | 20.31% | 8.73% |
| Jan-15 | 37.52% | 18.29% | 14.86% | 21.52% | 7.81% |
| Feb-15 | 37.57% | 13.85% | 14.99% | 25.24% | 8.35% |
| Mar-15 | 34.68% | 21.10% | 14.65% | 21.77% | 7.80% |
| Apr-15 | 32.63% | 25.68% | 15.26% | 18.73% | 7.70% |
| May-15 | 28.31% | 31.19% | 16.46% | 15.88% | 8.17% |
| Jun-15 | 24.26% | 39.92% | 14.81% | 14.81% | 6.21% |
| Jul-15 | 25.33% | 36.02% | 15.52% | 15.08% | 8.05% |
| Aug-15 | 30.08% | 26.74% | 18.32% | 16.44% | 8.42% |
| Sep-15 | 31.32% | 19.87% | 14.94% | 17.33% | 16.53% |
| Oct-15 | 34.39% | 23.45% | 14.58% | 19.02% | 8.56% |
| Nov-15 | 34.57% | 22.59% | 14.33% | 18.87% | 9.64% |
| Dec-15 | 32.91% | 26.76% | 14.47% | 17.36% | 8.50% |

## Appendix B. Interest by State

Table B1 consists of the interest in each of the terms in the States in descending order.

Table B1. States in declining order of interest for the five examined terms.

| Asthma | Lyme disease | Melanoma | COPD | Salmonella |
|---|---|---|---|---|
| Maine | Maine | West Virginia | West Virginia | Idaho |
| Delaware | New Hampshire | Tennessee | Kentucky | South Dakota |
| Kentucky | Vermont | Pennsylvania | Arkansas | North Dakota |
| Connecticut | Rhode Island | Vermont | Tennessee | Montana |
| Tennessee | Connecticut | Maine | Maine | Utah |
| West Virginia | Massachusetts | North Dakota | South Dakota | Wyoming |
| North Carolina | Pennsylvania | Alabama | Alabama | Arkansas |
| Mississippi | Maryland | Delaware | Wyoming | New Mexico |
| Rhode Island | Delaware | New Hampshire | Mississippi | Alaska |
| Maryland | New York | Massachusetts | Indiana | Vermont |
| Pennsylvania | New Jersey | Idaho | Ohio | Washington |
| Wyoming | West Virginia | Kentucky | North Dakota | Minnesota |
| Idaho | Wisconsin | Missouri | Missouri | Wisconsin |
| South Dakota | Virginia | Connecticut | Vermont | Mississippi |
| Nebraska | Iowa | Iowa | New Mexico | California |

| | | | | |
|---|---|---|---|---|
| Arkansas | Minnesota | Ohio | Oklahoma | Arizona |
| Alabama | Tennessee | Maryland | Montana | Nevada |
| Alaska | Indiana | Indiana | New Hampshire | North Carolina |
| Georgia | Kentucky | Arkansas | Pennsylvania | Alabama |
| South Carolina | North Carolina | Montana | Nebraska | Connecticut |
| Indiana | Arkansas | Rhode Island | Delaware | Iowa |
| Colorado | North Dakota | South Dakota | Michigan | Maryland |
| New Mexico | Michigan | Wyoming | South Carolina | Delaware |
| Missouri | Missouri | South Carolina | Iowa | Colorado |
| Oklahoma | South Dakota | Minnesota | North Carolina | Nebraska |
| Michigan | Alabama | Nebraska | Louisiana | New Hampshire |
| Vermont | South Carolina | North Carolina | Florida | South Carolina |
| Arizona | Montana | Mississippi | Minnesota | Hawaii |
| Montana | Illinois | New Jersey | Wisconsin | Oklahoma |
| New Jersey | Ohio | Florida | Connecticut | Rhode Island |
| Massachusetts | Kansas | Utah | Rhode Island | Maine |
| Minnesota | Mississippi | Wisconsin | Arizona | Georgia |
| Washington | Oklahoma | Michigan | Maryland | Missouri |
| Ohio | Nebraska | Arizona | Idaho | Indiana |
| Texas | Idaho | Oklahoma | Georgia | Louisiana |
| New York | Georgia | Colorado | Kansas | Massachusetts |
| Illinois | Florida | Georgia | Illinois | Kentucky |
| North Dakota | Wyoming | Kansas | Nevada | West Virginia |
| New Hampshire | Louisiana | Louisiana | Massachusetts | Ohio |
| Kansas | California | New Mexico | Alaska | Pennsylvania |
| Utah | Washington | Illinois | Colorado | Kansas |
| Wisconsin | Colorado | Washington | Texas | Texas |
| Florida | New Mexico | New York | New Jersey | Florida |
| Nevada | Utah | Hawaii | Washington | Tennessee |
| California | Alaska | California | New York | Illinois |
| Iowa | Arizona | Alaska | Hawaii | Michigan |
| Hawaii | Nevada | Texas | Utah | New Jersey |
| Louisiana | Texas | Nevada | California | New York |
| Virginia | Hawaii | Virginia | Virginia | Virginia |
| Oregon | Oregon | Oregon | Oregon | Oregon |

## References

[1] R. Kitchin, The real-time city? Big data and smart urbanism, GeoJournal. 79:1 (2014) 1–14.
[2] http://www.smart-cities.eu (accessed on May 5, 2016).
[3] D. Li, Y. Yao, Z. Shao, Big data in smart city, Geo. Spat. Inf. Sci. 39:6 (2014) 631–640.
[4] Z. Khan, A. Anjum, K. Soomro, M.A. Tahir, Towards cloud based big data analytics for smart future cities, J. Cloud Computing. 4:1 (2015) 11p.
[5] A. Holzinger, C. Röcker, M. Ziefle, From smart health to smart hospitals, Lecture Notes in Comput. Sci. 8700 (2015) 20p.
[6] H. Jung, K. Chung, Sequential pattern profiling based bio-detection for smart health service, Cluster Comput. 18:1 (2015) 209–219.
[7] A. Solanas, C. Patsakis, M. Conti, I. Vlachos, V. Ramos, F. Falcone, O. Postolache, P. Perez-Martinez, R. Pietro, D. Perrea, A. Martinez-Balleste, Smart health: A context-aware health paradigm within smart cities, IEEE Commun. Mag. 52:8 (2014) 74–81.
[8] C. Dobre, F. Xhafa, Intelligent services for Big data science, Future Gener. Comp. Sy. 37 (2014) 267–281.
[9] D.W. Bates, S. Saria, L. Ohno-Machado, A. Shah, G. Escobar, Big data in health care: Using analytics to identify and manage high-risk and high-cost patients, Health Aff. (Millwood) 33:7 (2014) 1123–1131.
[10] T. Braun, U. Harréus, Medical nowcasting using Google trends: Application in otolaryngology, Eur. Arch. Otorhinolaryngol. 270:7 (2013) 2157–2160.
[11] S.V. Nuti, B. Wayda, I. Ranasinghe, S. Wang, R.P. Dreyer, S.I. Chen, K. Murugiah, The use of google trends in health care research: A systematic review, PLoS ONE. 9:10 (2014) e109583.
[12] N.L. Bragazzi, S. Bacigaluppi, C. Robba, R. Nardone, E. Trinka, F. Brigo, Infodemiology of status epilepticus: A systematic validation of the Google Trends-based search queries, Epilepsy Behav. 55 (2016) 120–123.
[13] A.K. Johnson, S.D. Mehta, A comparison of internet search trends and sexually transmitted infection rates using google trends, Sex. Transm. Dis. 41:1 (2014) 61–63.
[14] C. Alicino, N.L. Bragazzi, V. Faccio, D. Amicizia, D. Panatto, R. Gasparini, G. Icardi, A. Orsi, Assessing Ebola-related web search behaviour: Insights and implications from an analytical study of Google Trends-based query volumes, Infect. Dis. Poverty. 4:1 (2015) 54.

[15] S. Pollett, N. Wood, W.J. Boscardin, H. Bengtsson, S. Schwarcz, K. Harriman, K. Winter, G. Rutherford, Validating the Use of Google Trends to Enhance Pertussis Surveillance in California, PLoS Curr. (2015) Edition 1.
[16] www.google.com/trends (accessed on April 29, 2016).
[17] http://www.cdc.gov/lyme/stats/maps.html (accessed on May 2, 2016).
[18] http://www.cdc.gov/lyme/stats/ (accessed on May 2, 2016).
[19] http://www.cdc.gov/copd/data2.html (accessed on August 9, 2016).
[20] https://nccd.cdc.gov/DCPC_INCA/ (accessed on August 9, 2016).
[21] http://www.aafa.org/page/asthma-capitals.aspx (accessed on August 9, 2016).
[22] http://www.cdc.gov/salmonella/reportspubs/salmonella-atlas/serotyping-importance.html (accessed on August 9, 2016).
[23] http://www.cdc.gov/salmonella/reportspubs/salmonella-atlas/serotype-snapshots.html (accessed on August 9, 2016).
[24] http://www.cdc.gov/lyme/stats/ (accessed on August 9, 2016).