

**Analyzing Risk Response Dynamics on the Web: the Case of
Hurricane Katrina**

Ka Lok Lee

*Wharton School
University of Pennsylvania*

Robert J. Meyer

*The Wharton School
University of Pennsylvania*

Eric T. Bradlow

*The Wharton School
University of Pennsylvania*

Working Paper # 2008-09-15

Risk Management and Decision Processes Center
The Wharton School, University of Pennsylvania
3730 Walnut Street, Jon Huntsman Hall, Suite 500
Philadelphia, PA, 19104
USA
Phone: 215-898-4589
Fax: 215-573-2130
<http://opim.wharton.upenn.edu/risk/>

CITATION AND REPRODUCTION

This document appears as Working Paper of the Wharton Risk Management and Decision Processes Center, The Wharton School of the University of Pennsylvania. Comments are welcome and may be directed to the authors.

This paper may be cited as: Ka Lok Lee, Robert J. Meyer and Eric T. Bradlow, "Analyzing Risk Response Dynamics on the Web: the Case of Hurricane Katrina," Risk Management and Decision Processes Center, The Wharton School of the University of Pennsylvania, December 2008.

The views expressed in this paper are those of the author and publication does not imply their endorsement by the Wharton Risk Center and the University of Pennsylvania. This paper may be reproduced for personal and classroom use. Any other reproduction is not permitted without written permission of the authors.

THE WHARTON RISK MANAGEMENT AND DECISION PROCESSES CENTER

Since its creation 25 years ago, the mission of the Wharton Risk Management and Decision Processes Center has been to carry out a program of basic and applied research to promote effective corporate and public policies for low-probability events with potentially catastrophic consequences. The Risk Center has focused on natural and technological hazards through the integration of risk assessment, risk perception and risk financing with risk management strategies. After 9/11, research activities extended also to national and international security issues (e.g., terrorism risk insurance markets, protection of critical infrastructure, global security).

Building on the disciplines of economics, decision sciences, finance, insurance, marketing and psychology, the Center's research program has been oriented around descriptive and prescriptive analyses. Descriptive research focuses on how individuals and organizations interact and make decisions regarding the management of risk under existing institutional arrangements. Prescriptive analyses propose ways that individuals and organizations, both private and governmental, can make better decisions regarding risk. The Center supports and undertakes field and experimental studies of risk and uncertainty to better understand the linkage between descriptive and prescriptive approaches under various regulatory and market conditions. Risk Center research investigates the effectiveness of strategies such as risk communication, information sharing, incentive systems, insurance and regulation.

The Center is also concerned with training decision makers and actively engaging multiple viewpoints, including the expertise of top-level representatives in the world from industry, government, international organizations, interest groups and academics through its research and policy publications and through sponsored seminars, roundtables and forums.

More information is available at <http://opim.wharton.upenn.edu/risk>

Analyzing Risk Response Dynamics on the Web: the Case of Hurricane Katrina

Ka Lok Lee
Robert J. Meyer
Eric T. Bradlow*

December 2008

* Ka Lok Lee is a doctoral candidate in Marketing, Robert J. Meyer is Gayfryd Steinberg Professor, Professor of Marketing and Co-Director, Wharton Risk Management and Decision Processes Center, and Eric T. Bradlow is the K.P. Chao Professor, Professor of Marketing, Statistics and Education, and Co-Director of the Wharton Interactive Media Initiative, all at the Wharton School of the University of Pennsylvania. All correspondence on this manuscript should be sent to Ka Lok Lee, Wharton School Marketing Department, 3730 Walnut Street, Suite 700 JMHH, Philadelphia, PA 19104; kaloklee@wharton.upenn.edu. We want to thank Leonard Lodish from Wharton for suggesting the original analysis and providing the contact with Compete.com. We also want to thank Sarah Ratcliffe from the Department of Biostatistics at Penn for her useful feedback on the first draft of the paper.

Abstract

Using a unique dataset that documented the hourly web-surfing behavior of over 140,000 Internet users in five southeastern states in August of 2005, we explore the dynamics of information gathering as Hurricane Katrina developed and then hit South Florida and the Northern Gulf Coast. Using both elementary statistical methods and advanced techniques from functional data analysis (Ramsay and Silverman 2005), we examine both how storm events (such as the posting of warnings) affected traffic to weather-related web sites, and how this traffic varied across locations and by characteristics of the web user. A general finding is that spatial-temporal variation in weather-site web traffic generally tracked the timing and scale of the storm threat experienced by a given area. There was, however, considerable variation in this responsiveness. Residents in Florida counties that had been most directly affected by Hurricane Dennis just a month earlier, for example, displayed more active visitation rates than those who had been less affected. We also find evidence of a gender effect where male users displayed a disproportionately larger rate of visitation to weather sites given the onset of storm warnings than females. The implications of this work for the broader study of behavioral response dynamics during hazards are explored.

Key words: Risk Perception, Risk Communication, Clickstream Data, Exploratory Data Analysis, Functional Data Analysis

1. Introduction

It is estimated that almost 70% of the U.S. population are Internet users¹. The World Wide Web has become an important medium for the buying and selling of goods (e.g., www.amazon.com), the gathering of information (e.g., www.google.com), a place for people to “socialize” (e.g., www.facebook.com), or perhaps just get their “15 minutes” of fame (e.g., www.youtube.com). However, the web also plays a far less frivolous role as a major channel for the dissemination of news information that can be critical during crises or natural disasters. When the *New Orleans Times Picayune*’s printing presses were disabled by Hurricane Katrina in 2005, for example, residents were able to get uninterrupted news about the emerging disaster and relief effort through its companion website, www.NOLA.com. Likewise, the recent rapid growth of visits to weather-related websites such as www.weather.com and www.nhc.noaa.gov suggests that the web is rapidly augmenting—if not replacing—traditional media such as radio, television, and newspapers as the method by which individuals learn about threats from natural hazards.

The degree to which the web is effective as a means of informing the public about natural hazards, however, is far from clear. On the one hand, the ability of the web to store and transmit large volumes of information means that it can provide far more timely and detailed information about hazards than would ever be possible using traditional media such as radio and television (e.g., Bucher 2002; Dunbar 2007). On the other hand, this enhanced information is of value only to the degree that it is actively sought and utilized in a timely manner by residents before and after a hazardous event. While

¹ Pew Internet & American Life Project <<http://www.pewinternet.org/trends.asp>>

survey-based research that suggests that residents are making greater use of the internet as a source of information about hazards (e.g., Bucher 2002; Piotrowski and Armstrong 1998), there has been little direct evidence on the dynamics of how and when hazard-related web sites are accessed and used during the course of an actual natural disaster.

The purpose of this paper is to report such an analysis. Using a unique dataset containing the complete web-surfing habits of over 140,000 web users in the southeastern United States, we describe how internet usage evolved before, during, and after a particularly notable recent event, Hurricane Katrina. In this analysis we relate spatial and temporal variation in the frequency of visitation to weather-related information sites to events that would have differentially affected the risk faced by residents, such as the timing and location of warnings issued by the National Hurricane Center.

We see this analysis as making three sets of contributions. First, at a conceptual level, it illustrates how web-surfing patterns can be used to provide insights into an aspect of risk response that has traditionally gone unexplored: how and when environmental threats trigger information gathering in advance and after exposure to a hazard. Second, it makes a substantive contribution by using web-usage data to test a series of basic hypotheses about these dynamics in the context of hurricanes, namely how levels of information gathering for weather information co-varies with such factors as objective storm threat and the gender, age, and past storm experience of the web user. Finally, it makes a methodological contribution by demonstrating how the enormous complexities that come with using web data to test such hypotheses can be overcome using a relatively new approach to exploratory data analysis, functional data analysis (FDA).

The remainder of this paper is as follows. In Section 2, we describe the data used in the research. In Section 3, we report an exploratory analysis that illustrates some of the basic features of web usage before, during, and after the storm at different locations. We report a more systematic analysis of the spatial and temporal drivers of web usage in Section 4. We conclude in Section 5 with a summary of the findings and a discussion of its implications for the use of web data as a tool for understanding individual response to hazardous events.

2. The Data

Our clickstream data described the web-surfing behavior of 141,100 individual web users living in the states of Alabama, Florida, Louisiana, Mississippi, or Tennessee. The data were provided by *Compete.com* (www.compete.com), a firm that tracks and analyzes consumer web usage. While our panel of *Compete.com* users should not be considered an entirely random sample, it is a very large and informative self-selected convenience sample. The data spans the period from August 15th, 2005 to August 31st, 2005, coinciding with the before, during, and after lifespan of Hurricane Katrina. Every URL viewed by a user is recorded with a timestamp and a date. The data also contain individual-user information on gender, age (categorized into 7 groups²), income (categorized into 4 groups³), and zip codes of the users. Within the sample 45% of panelists are female, 26% are younger than 34, 63% are between 35 and 64, and 11% are 65 or above. This demographic profile is comparable to that estimated for more general internet-user populations by other sources (e.g., Pew Internet Life Project 2008). In total, the dataset has more than 100 million observations.

² Age is categorized into “<18”, “18-24”, “25-34”, “35-44”, “45-54”, “55-64”, and “65+”.

³ Income is categorized into “<30K”, “30K-60K”, “60K-100K”, and “100K+”.

Because of the enormous scope of the data, to achieve initial parsimony we first assigned each site visited by each panelist to one of sixteen *a priori* defined content categories (such as news, weather, recreation, etc.) using site categorization software developed by www.lightspeedsoftware.com. We then aggregated the data to the hourly frequency of usage by panelists within geographic counties. This resulted in a final data set of 408 hourly measurements of visitation frequencies in 375 counties in 5 states to sixteen site categories. Of central interest in this research is visitation to sites in the *weather* category, which consists of all the websites providing primarily weather-related information that were ever visited by our panelists in the study period.⁴ While residents likely obtained information from news sites (and presumably many did), the multi-functional nature of such sites clouded the interpretability of temporal visitation patterns. Specifically, exploratory analyses showed that while visitation of weather sites closely tracked events during the course of the storm (such as the issuance of warnings), this was less the case for usage of news sites.

The particular response measure we tracked was the proportion of unique users visiting weather websites in a county within a given hour. Formally, if N_c is the total number of users in county c and $W_c(t)$ is the number of unique users visiting weather websites in county c between the hour $t-1$ and the hour t , the proportion of unique users visiting weather websites in county c between the hour $t-1$ and t is given by the ratio $\%W_c(t) = W_c(t)/N_c$. An preliminary analysis of the 375 counties revealed that some were marked by extremely limited levels of web usage due to small population sizes, raising concerns about the stability of some county-level estimates of $\%W_c(t)$. As a result, we

⁴ We went through an exhaustive process with our data provider and IT analyst to ensure that we included all the relevant websites. Some of the websites in the weather category are www.weather.com, www.weatherunderground.com, and the National Hurricane Center's site (www.nhc.noaa.gov).

focused all subsequent analyses on those counties with at least 60 users (based upon a median split of county user sizes). This reduced the number of counties to 187 with 136,484 users remaining.⁵

3. Hurricane Katrina and general features of web usage

In many ways, Hurricane Katrina offered a unique laboratory for the study of spatial-temporal variation in response to hazards. It was one of only a small number of hurricanes in history to directly strike two major urban areas (Miami-Dade County, Florida and New Orleans), with each strike having a different character: when the storm struck Miami-Dade it had formed just the day before giving residents rather limited advance notice, and was of marginal (category 1) strength upon landfall. In contrast, prior to hitting the Northern Gulf Coast Katrina had intensified into one of the largest and strongest Atlantic hurricanes on record, and, of course, caused unprecedented damage upon impact. Likewise, the areas that it affected were marked by a diversity of recent experiences with hurricanes. Whereas Southeast Florida had not been directly affected by a hurricane in over thirteen years (Andrew in 1992), the Florida panhandle had been hit by a small but powerful hurricane (Dennis) just one month before Katrina, and New Orleans had experienced a major false alarm the previous September when the city was evacuated in advance of Hurricane Ivan (which ultimately hit the Florida/Alabama border area).

Because a large amount of information on Hurricane Katrina is available elsewhere (e.g., Knobb, Rhome and Brown 2005), we only reiterate selected information about its history relevant to our event analysis. At 11AM on August 24th, a tropical

⁵ Note that even though we have eliminated 50% of the counties, this only eliminates 4,516 users (less than 4%).

depression in the southeast Bahamas was upgraded to a Tropical Storm and named “Katrina”. At the same time a hurricane watch⁶ was issued for southeast Florida. At 11PM, the watch was upgraded to a hurricane warning⁷. At 5PM on August 25th, Katrina was upgraded to a Category-1 hurricane, and at 7PM made landfall on the Dade-Broward, FL county line. After moving into the southeastern Gulf of Mexico, at 11AM on August 27th a hurricane watch was issued for the north-central Gulf Coast, including the city of New Orleans. At 6PM, New Orleans Mayor C. Ray Nagin declared a state of emergency and issued a mandatory evacuation order. At 11PM, the hurricane watch was upgraded to a hurricane warning. At 7:10AM on August 29th Katrina made its second landfall near Buras-Triumph, Louisiana and passed just east of New Orleans at 10AM. Once over land Katrina began to weaken, and by 8PM, Katrina was downgraded to a tropical storm over central Mississippi. The storm ceased being a traceable entity on August 30th while moving through central Tennessee.

To illustrate the large amount of spatial and temporal variability that existed in web users’ responses to these events, in Figures 1(a) through 1(c) we plot observed geographic variation in mean daily relative frequencies of weather-site visitation for each county at each of three points in time: August 19th, four days before the storm is formed (a representative baseline), August 24th, when the storm formed and warnings were issued for South Florida, and August 27th, when the storm was in the Gulf and warnings were issued for the northern Gulf coast. In each of the figures, a darker color corresponds to higher relative frequencies of access to weather sites. The maps effectively convey the

⁶ A hurricane watch means that hurricane conditions are possible within the watch area, generally within 36 hours (NOAA National Hurricane Center).

⁷ A hurricane warning means that hurricane conditions are expected within the warning area with the next 24 hours (NOAA National Hurricane Center).

central challenge in analyzing the data: while they suggest some tendency for visitation rates to respond in an intuitive manner to storm events, interest in the storm varied greatly over the geographic under study, and not in a way that was perfectly correlated with the threat faced by individual counties. When Katrina first formed and warnings were first issued for South Florida on the 24th, for example, Figure 1(b) shows an increase in weather-site traffic not just in the primary warning area (Dade, Broward, and Palm Beach Counties), but throughout the entire Florida peninsula and the distant panhandle. In contrast, we see little evidence of increased interest in the northern Gulf Coast.

INSERT FIGURE 1 HERE

To provide a more precise look at how weather-site traffic varied over time in response to storm events, in Figures 2a and 2b we plot the frequency of visitation to weather web sites for two illustrative counties that were directly affected by the storm: Miami-Dade, Florida (Figure 2a) and New Orleans, Louisiana⁸ (Figure 2b). The plots suggest these visitation patterns were marked by four common features:

1. High diurnal variation in visitation rates corresponding to daily sleep/awake cycles;
2. Increases in visitation rates that generally correspond to the timing of the release of warnings;
3. An “immediate threat” effect, where visitation rates were greatest when the location was subject to an explicit warning; and
4. An “aftershock” effect, where visitation rates remained well-above pre-storm levels after the storm had passed.

⁸ The city of New Orleans is coextensive with Orleans Parish, meaning that the boundaries of the city and the parish (i.e., county) are the same.

INSERT FIGURE 2 HERE

As an illustration, in the first 8 days of the data panelists in Miami-Dade displayed a stable baseline rate of visitation to weather web sites, marked by a daily maximum that occurred between 5PM and 7PM and daily minimum occurring between midnight and 6AM. There is a slight increase in traffic on the 23rd, coinciding with issuance of the first advisory on a tropical depression that would later become Katrina, and then a dramatic increase on the 24th coinciding with the issuance of the first explicit *hurricane watch* at 11AM. Traffic volume then continued to increase through the day, reaching a peak at the time of the usual diurnal maximum at a level approximately 7 times more than the baseline. On the 25th, when the area was under a hurricane warning, usage again rapidly rose during the day, reaching a peak near the time of the storm's landfall, which corresponded to the time of the natural diurnal maximum of web use. The temporal pattern in New Orleans was similar to that observed in Miami-Dade, with visitation rates closely co-varying with the level of threat faced by the city at a given point in time. For example, traffic remained near baseline levels when Katrina was hitting South Florida on the 25th, but increased on the 26th when the storm entered the southeastern Gulf of Mexico and residents were advised to monitor future bulletins. It then peaked on the 27th when a hurricane warning was issued. Visitation rates decreased on the 28th in response to the evacuation order that was issued the previous evening.⁹

Both plots also show a more subtle “aftershock” feature, where mean visitation levels in both Miami and New Orleans remained above pre-storm levels after the storm

⁹ It is curious to note that despite the mandatory evacuation, web usage in New Orleans remained quite high both during and after the storm. We assume that this reflects a blend of usage by residents who did not evacuate and had continued access to power sources or, more likely residents who evacuated with their computers/laptops, and were accessing the web from remote locations.

had passed. A natural explanation for this heightened post-storm traffic is that residents' direct experiences with Katrina inflated (if only temporarily) their worry about being affected by future storms, and/or they became more familiar with the sites by virtue of their use during the storm (a bookmarking effect).¹⁰

While the web-visitation patterns in these two cities showed communalities, it is important to note that they also displayed important differences. For example, web users in Miami-Dade acted as though they were more “storm wary” than those in New Orleans by responding more quickly to the release of storm information. The release of the first advisory indicating that a depression had formed in the Bahamas on the 23rd—a weak warning signal—was associated with almost a doubling of visits to weather sites in Miami-Dade compared to pre-storm levels. In contrast, the much stronger signal of Katrina as a full hurricane hitting Florida on the 25th and headed for Gulf induced only a 40% increase in traffic in New Orleans. In the next section we undertake a more systematic analysis of the data that attempts to explain such observed variance in spatial-temporal response patterns.

4. Explaining Variation in Weather Site Usage

In this section we undertake a more systematic approach to understanding the drivers of individual differences in weather-site visitation patterns over time and space during Katrina. While there has been quite a bit of prior research that has sought to explain both individual differences perceptions of risk from hazards (e.g., Fischhoff et al. 2005; Gustafson 1998; Peacock, Brody, and Highfield 2005; Perry, Lindell, and Green 1982) and protective actions in response to risk (e.g., Peacock 2003), the degree to which

¹⁰ It is unlikely that these increased visitation levels reflected attempts by users to learn about the damage aftermath of the storm, as the weather sites under study tended to carry only information about contemporaneous weather conditions and forecasts, not news stories about the storms *per se*.

such findings have a direct mapping to how the web is used during a hurricane hazard is far from clear. In the sections below, we first take up this question by analyzing the nature of geographic differences in both the timing and magnitude of web response to the storm, and then turn to the effects of gender, age and income.

Storm impact by time and region

As noted above, Hurricane Katrina offered something of a natural experiment for the study of the drivers of geographical variation in web-based information gathering in response to a storm threat. In addition to having two independently threatened zones (South Florida and the northern Gulf), the fact that the areas differed in their historical experiences with hurricanes provided a means for a large-scale field test of hypotheses about memory-decay effects on hazard response. Specifically, a good bit of prior work has demonstrated that perceived vulnerability to hazards tends to be influenced by how recently they were last directly experienced. Hence, demand for flood insurance decreases as the years since a last flood increases (Kunreuther 1983), and American's judgments about the likelihood of being hit by a terrorist attack has decreased in the years after the 9/11 attack (Fischhoff, et al 2005). We might hypothesize, therefore, that the propensity of residents to quickly turn to the web to gather storm information would be highest among those who have most recently experienced a hurricane in our data set (e.g., the Florida Panhandle, which had experienced Hurricane Dennis just a month earlier), and lowest among those in inland locations (such as Tennessee) who would have had little experience with such storms.

What makes this hypothesis somewhat less transparent, however, is that previous work is largely silent about how *rapidly* risk perceptions subside after a hurricane

experience, or how severe the initial storm experience has to be to induce a sense of heightened vulnerability. As noted above, the previous (2004) hurricane season had been an extremely active one in which storms affected almost the entire study area—including inland locations—to one degree or another. Hence, if the 2004 season had induced a heightened global awareness of hurricanes, we may see limited variations in responsiveness across areas.

Data preparation and smoothing using functional analysis. Because our central interest was in testing hypotheses about how levels of actual risk affected web utilization during the storm, to prepare the data for analysis we first categorized counties into eight “risk zones” that were relatively homogeneous in terms of the level and timing of the threat posed by Katrina (see Figure 3). These zones were:

1. South Florida (under a hurricane warning on August 25th)
2. The north-central Gulf Coast (under a hurricane warning from August 27th-29th)
3. The Florida Panhandle (under a hurricane watch on August 28th)
4. Inland Mississippi and western Alabama (under inland high wind warnings on August 29th)
5. Tennessee (no warnings)
6. Central and North Florida (no warnings)
7. Inland eastern Alabama (no warnings)
8. The western Louisiana Gulf Coast (no warnings)

INSERT FIGURE 3 HERE

As foreshadowed by the above exploratory analysis, analyzing the spatial-temporal pattern of the data posed us with the challenge of how to control for the large amount of natural variation that marked the day-to-day visitation of web sites—both due to sleep/awake cycles and random influences on web use (such as proximity to a computer). To handle this variability we turned to the methods of *functional data analysis* (FDA), a relatively new approach to analyzing variation across multiple time series (Ramsay and Silverman 2005).¹¹ The central idea behind FDA is that each sequence of time series data (in our case, hourly web usage over time for each individual county) is assumed to be a “sample”, possibly measured with random error, from a smooth function of time. In order to turn a sequence of discrete data into a function, the analyst first seeks to identify the parametric form best describing the data. Then she can use the functional counterparts of traditional statistical analysis methods (such as analysis of variance or principal components analysis) to explain variations across a set of functions.

The temporal variation in the percentage of weather-site users in a particular county over time ($\%W_c(t)$) could be represented by a basis expansion:

$$\%W_c(t) = \sum_{k=1}^K a_k \phi_k(t), \quad (1)$$

where $\phi_k(t)$ are linearly independent and known functions, called basis functions, and a_k are coefficients that link each basis function together in the representation of $\%W_c(t)$. While each sequence of discrete data is represented by the same basis expansion, the set of coefficients, a_k , is different. The most obvious feature of web visitation over time

¹¹ All functional data analyses in this paper were performed using the publicly available MATLAB toolbox obtained from www.functionaldata.org, a website managed by Professors Ramsay and Silverman.

that we noted in the raw data was the daily sinusoidal pattern corresponding to sleep/wake periods. To reflect this pattern we chose for $\phi_k(t)$ the Fourier system: $\phi_k(t) = \{1, \sin(\omega t), \cos(\omega t), \sin(2\omega t), \cos(2\omega t), \sin(3\omega t), \cos(3\omega t), \dots\}$, where the parameter ω controls the periodicity of the variation $= 2\pi/\omega$.

In FDA the dimensionality of the expansion K , i.e. how many terms to include in equation (1) to get an accurate representation of the data, is treated as a user-specified integer parameter. If K is chosen too small, the basis expansion gives a very smooth function, which may miss some key characteristics of the data. For example, in our case, a low K would miss very important peak events. But if K is chosen too large, we may end up fitting noise. For this application, we chose $K= 39$ from a range of tested values as the one that offered the best compromise between allowing a fairly smooth representation of the raw data while capturing the most important time-varying characteristics, such as peak usage given hurricane warnings.¹²

Results. We analyzed the degree to which the eight risk zones systematically differed in the temporal pattern of web usage they exhibited during Katrina using *functional analysis of variance* (fANOVA). As suggested by its name, fANOVA is similar to classical ANOVA in that we regress a response on a design matrix.¹³ However, now the responses are smoothed functions over time, and the regression coefficients also become coefficient functions over time.

Formally, let $\%W_{cs}(t)$ represent the proportion of weather website visitors for county c which is nested within zone s . The fANOVA model is:

$$\%W_{cs}(t) = \mu(t) + \alpha_s(t) + \varepsilon_{cs}(t) \quad (2)$$

¹² Analyses with other values of K are available upon request.

¹³ Our design matrix is a matrix with 1's and 0's in appropriate positions reflecting membership in a risk zone (or not).

where $\mu(t)$ is the grand mean function, which indicates the average proportion across all 8 zones over time, $\alpha_s(t)$ are zone specific effects reflecting higher or lower zone-level web usage over time, and $\varepsilon_{cs}(t)$ is the unexplained variation specific to county c in zone s . For identification, we constrain that the sum of all zone effects is zero. We can see that if we take away the time notation, (t) , the fANOVA model is the same as a classic ANOVA model. Details of our specific application of the fANOVA model are given in the appendix. Before we performed fANOVA, we processed the original data by subtracting the baseline level of visitation to weather websites prior to the impact of Katrina so that we could more easily examine how the visitation pattern *increased* due to Katrina.

In Figure 4 we plot the grand mean regression function, $\mu(t)$, with its corresponding 95% confidence interval. By visual inspection it is evident that inflections in the function closely track major changes in warnings and effects of the storm. On the 10th day, i.e., Aug. 24th, for example, the grand mean function starts to rise above zero (the baseline), corresponding to the timing of the issuance of the first hurricane watch by the National Hurricane Center. The peak of the grand mean function then coincides with the early afternoon of the day before Katrina made landfall on the Gulf Coast.

INSERT FIGURE 4 HERE

Of central interest in our analysis, however, is how usage patterns differed among the various risk zones. In Figure 5 we decompose the grand mean function into its eight zone-specific components, akin to plotting the marginal means of an experimental factor in a more traditional ANOVA, and in Figure 6 we plot the corresponding F-ratio function. Like the F-test in a traditional ANOVA, the F-function is an omnibus test of the degree to which the differences in marginal usage functions plotted in Figure 5 could be attributed

to chance variation in the data. Congruent with an intuitive read of Figure 5, the F-ratio function cannot reject a null hypothesis that prior to August 24th all zones exhibited daily patterns of visitation to the weather web sites, but began to differ significantly starting on the 24th, when each zone began to be differentially threatened by the storm.

INSERT FIGURES 5 & 6 HERE

For the most part the figures tell a simple story about web response to the storm illustrating the nature of these differences:

1. The temporal maxima of weather-site traffic for each zone roughly corresponds to the timing of the issuance of warnings and/or the time of the storm's closest approach, and the size of the maximum roughly corresponds to the magnitude of the threat experienced by a zone;
2. There is a sustained interest (or aftershock) effect, where visitation rates after the storm had passed remained higher than base rates before it.

To demonstrate, as soon as warnings were posted on the 24th, traffic in South Florida immediately increased, and this was followed by a delayed—and somewhat more subdued—increase in traffic among residents in central Florida, who were just outside the warning zone. Then, even after the storm had moved inland over the Gulf, rates of visitation in both regions remained higher than pre-storm baselines (an effect, as noted above, that could be reflective of heightened wariness over the possibility of additional future storms). In contrast, the zone that shows that last upswing in traffic was Tennessee, which was, of course, the last to be threatened by the storm.

Where this pattern does *not* clearly hold, however, is the temporal pattern for the Florida Panhandle. Web traffic showed an increase in the Panhandle on the 24th even

though the zone was some distance from South Florida and under no immediate threat, and users in the Panhandle showed the highest overall rate of visitation to weather sites during the course of the storm. In contrast, we see no such early-interest among residents along the Alabama, Mississippi, and Louisiana coasts. One possible explanation for this early increase in visitation is that the initial 5-day forecast path produced by the National Hurricane center on the 24th projected the storm to cross the Florida Peninsula on the 25th and then make landfall on the Panhandle 2-3 days later (see http://www.nhc.noaa.gov/archive/2005/KATRINA_graphics.shtml). While the certainty of this forecast was quite low (for example, the forecast odds of landfall within 65 miles of either Panama City and Pensacola at that time were less than 10%), intense media coverage may have induced residents to take the storm threat much more seriously than forecast odds might have prescribed (see, e.g., Broad, Leiserowitz, Weinkle, and Steketee 2007)¹⁴.

There is, however, another possible explanation: at the time Katrina formed residents of the Florida Panhandle would have still been recovering from the damage caused by Hurricane Dennis (a much smaller storm) which had hit the area the month before. To explore whether the impact of Dennis might have amplified the storm-wariness of Panhandle residents, in Figure 7 we plot the patterns of weather-site visitation for a set of four contiguous counties in the western Florida Panhandle and neighboring Alabama. The data provide an intriguing result: the interest in Katrina in the Panhandle shown in Figure 5 was, in fact, *not* universally shared by all counties, but was rather driven by active early responses in one: Santa Rosa, which was the county that was the point of landfall of Dennis and which experienced the highest wind speeds during its

¹⁴ We thank an anonymous referee for this suggestion.

passage (Beven 2005). Hence, a reasonable interpretation of the data is that Santa County residents' recent experiences with Dennis had put them particularly on edge with respect to potential storm threats, and this triggered active information gathering the moment they saw they were again in a projected path, even though the certainty of this path was low.

INSERT FIGURE 7 HERE

Gender, Age and Income Effects

Whether there would be gender, age, or income effects on web-usage responses to Katrina is far from certain. To illustrate, one of the most robust findings in prior studies of risk perception is that women tend to be more risk-sensitive than men, both by perceiving themselves as being more vulnerable to threats (e.g., Flynn, Slovic, and Mertz 1994; Gustafson 1998), and by more likely to take evasive actions in response to threats (e.g., Gustafson 1998). A simple extrapolation of these findings would thus suggest the hypothesis that women would be more likely to increase their rate of visitation to storm-related weather sites in response to Katrina (relative to baseline usage) because they would be more motivated to understand their risk exposure. On the other hand, one might also hypothesize that visitation rates among women would *decrease* during storms as their time would be spent engaging in more overt protective acts (such as stocking up on foods) rather than passively gathering information on the web. That is, while men monitor the storm on television and the web, women are actually taking the protective actions.

The effect of age on usage is similarly unclear. On the one hand, the fact that overall internet familiarity and usage tends to be higher among younger age groups (e.g.,

<http://www.pewinternet.org/trends.asp>) suggests the hypothesis that younger residents will show the greatest tendency (relative to baseline) to turn to the web to monitor storm information when a threat appears. In contrast, this familiarity effect may be offset by the fact that perceived vulnerability to hazards tends to be highest among older age groups (e.g., Fischhoff et al 2005).

Not surprisingly, income also has a bi-directional set of predictions (albeit probably not to the degree of age or gender). On the one hand, one could argue that persons of lower income are more threatened by the storm (and hence would have increased web usage) as property damage could be higher for less well-constructed properties and living areas with a less developed infrastructure. Oppositely, one could imagine that wealthier individuals utilize the web more often as an information source and/or “have more to lose”. While our expectations would probably lean towards an increased web usage positively correlated with income, the data will allow us to address this empirically.

To explore the effect of gender, in Figure 8 we plot two alternative ways of examining the effect of gender on web-site visitation patterns: the probability that a user would be visiting a weather web site given that they are either a man or a woman (visitation conditional on gender, Figure 8a) and the probability that a weather-site visitor was a man or a woman (gender conditional on visitation, Figure 8b). For ease of analysis we pool over all risk zones. Figure 8a shows two effects:

- 1) A constant and largely equal rate of visitation to weather sites for both men and women before the 24th, when the storm first formed and warnings were issued; and

2) Higher rates of visitation among men during the threat period after the 24th; Specifically, while both men and women increased their rate of visitation to weather web sites in response to the issuance of warnings, this increase was greater among men users compared to women.

INSERT FIGURE 8 HERE

To more rigorously test this prediction we partitioned the usage rates within gender into three time blocks: before the 24th, between the 24th and 29th (landfall), and after the 29th, and subjected the difference in visitation rates within each time block to a Wilcoxon Signed-Rank Test. Based on this we cannot reject a null hypothesis of equal visitation rates in the days *prior* to the first storm warnings ($p=.98$), but could reject it with strong evidence *afterward* ($p<.001$). However, Figure 8b shows that conditional on being a visitor, there is little change in the ratios of gender across time. This suggests a constant scaling up effect in the numbers for both genders (an interesting pattern).

In Figures 9 and 10, we show the visitation plot conditional on age and income, respectively. Unlike gender, here we do not see a systematic effect of age or income on weather-site visitation over time, and we cannot statistically reject a null hypothesis of equal visitation rates. Furthermore, while not shown, the conditional plots of age and income on visitation do not reveal any differential effects across time.

INSERT FIGURE 9 & 10 HERE

5. Discussion

Understanding how individuals use different communication media during the course of natural hazards is critical to attempts to improve emergency preparedness. Yet, despite the importance of such a study, to date we have had little understanding of the

dynamics of information utilization as hazards unfold, such as what triggers threatened residents to begin actively seeking information about an oncoming hazards, and what biases may exist in this information seeking. In this paper we took a step toward answering such questions for one channel of communication: information gathering on the World Wide Web. Using a unique Internet clickstream dataset, we performed a series of visual exploratory and more formal analyses to understand the changes in weather-site Internet usage as 2005's Hurricane Katrina formed, threatened, and then hit two heavily populated coastlines in the southeastern United States.

The results of this analysis were intriguing both substantively and theoretically. First, from a substantive perspective, the data support the casual observation that residents in threatened areas *do* turn to the web for detailed information as storms approach, and the timing and scale of traffic closely mirrors the threats faced by different locations. At the same time, the fact that large increases in traffic tended to occur only when specific areas came under threat suggests for many (or most) web users the weather sites serve as a *secondary* rather than primary channel for the warning dissemination; they turn to the web for details after first hearing that there is a need from other sources, such as word-of-mouth, television, or perhaps other web sources (like chats). Understanding how people utilize the web in conjunction with other information channels would be a future research opportunity when such data become available.

From a theoretical perspective the data provide interesting insights into the generalizability of previous findings on the drivers and dynamics of risk perceptions. Of particular interest was the finding that web-surfing behavior seeking storm information showed very clear recency effects, reinforcing previous findings that beliefs about

vulnerability to hazards are tied to how imaginable they are (e.g., Fischhoff et al. 2005). In all locations, the rates of visitation to weather web sites after the passage of Katrina remained above pre-storm levels and were the highest in the Florida-Panhandle counties that had received the most direct hit from Hurricane Dennis a month earlier. In contrast, the data give a somewhat different view of the effect of gender on information gathering than might have been predicted based on previous work. While previous work has found that women tend to be more likely to take preparedness actions in response to threats than men (e.g., Gustafson 1998), here we found it was the men who showed the greatest increase in visitation rates to weather sites when storm warnings were issued. A possible reconciliation that we noted is that women *were* being more responsive, but by engaging in more overt acts of preparation (such as stocking up on food) rather than seeking more information on the web.

At a broader level, we also see this paper as providing an illustration of how web-tracking data can be used to study behavioral phenomena that would be difficult to examine using more traditional survey-based methods—in this case the dynamics of information seeking during the time and course of a natural hazard. In that light, it is possible that even richer insights into hazard-response behavior could be obtained by looking at how panelists were using the web during Katrina while *not* gathering storm information on weather sites. For example, in Figure 11, we plot the overall traffic to FEMA and other governmental websites carrying disaster relief information. While the numbers are small relative to the sample size, it conveys the expected relationship with storm events: visitation rates rapidly increase at the very end of the time period, when the focus of residents shifted from storm preparation to storm recovery. Moreover, most of

the visitors were web users in Florida and Louisiana, the two states that receive direct impacts from the storm. While our data period ended before September 2005, having additional data on web visitation and other channels would be another future research opportunity to understand how affected residents sought relief from FEMA.

INSERT FIGURE 11 HERE

Finally, the analysis also served to illustrate the fact that the richness of the information carried by web data can sometimes act as a double-edged sword; the enormous size and complexity of web-tracking data can make the task of extracting any insights difficult. In this paper we showed how one, relatively new, approach to data exploration—functional data analysis—can be used to aid in this data-mining process. A natural next direction would be to further explore how to more rigorously conduct event analyses using FDA and other, possibly more structured, spatio-temporal modeling methods (e.g., Ma and Carlin, 2007).

Technical Appendix: fANOVA

We can represent the fANOVA model compactly in matrix notation:

$$\%W = Z\beta + \varepsilon \text{ with a constraint } L\beta = 0 \quad (\text{A1})$$

where $\%W$ is the functional response vector corresponding to the percentage of visitors to weather web sites, Z is a 187 x 9 design matrix, with the first column for the overall mean (or the intercept term) and the remaining 8 columns indicating zone-membership, β is 9-dimensional vector of regression functions, ε is the vector of residual functions, and L is a suitable matrix so that individual zone effects sum to zero.

We want to estimate β , the regression functions of the fANOVA model given in (A1). Furthermore, we want β to be smooth for easy interpretation. To achieve this goal, we express β with a basis expansion, as we do $\%W$, the functional response. This means that $\%W(t) = A\phi(t)$ and $\beta = B\theta(t)$ where $\phi(t)$ and $\theta(t)$ are basis functions of dimension K_y and K_β , respectively and A and B are linear coefficients. It is generally assumed that $\phi(t)$ and $\theta(t)$ follow the same basis system. Hence we utilize Fourier expansions for both. Then, the least squares criterion is:

$$\int [A\phi(t) - ZB\theta(t)]^T [A\phi(t) - ZB\theta(t)] dt \quad (\text{A2})$$

which can give us B and its confidence interval using the derivations given in Ramsay and Silverman (2005). For this application, K_y is chosen to be 39 as noted. K_β is chosen to be 19 so that we can have fairly smooth regression functions.

References:

Beven, Jack (2005), "Tropical Cyclone Report: Hurricane Dennis, 4-13 July 2005". Technical Report, NOAA National Hurricane Center.

Broad, Kenneth, Leiserowitz, Anthony, Weinkle, Jessica, and Marissa Steketee (2007), "Misinterpretations of the 'Cone of Uncertainty' in Florida During the 2004 Hurricane Season", *Bulletin of the American Meteorological Society* 88 (5), 561-667.

Bucher, Hans-Jeurgan (2002), Crisis Communication and the Internet: Risk and Trust in a Global Media, *First Monday*, 1-10.

Dholakia, Ruby R., Dholakia, Nikhilesh, and Nir Kshetri (2003), "Gender and Internet Usage", the *Internet Encyclopedia*

Dunbar, Paula K. (2007), "Increasing Public Awareness of Natural Hazards via the Internet," *Natural Hazards*, 2007, 42 (3), 529-536.

Fischhoff, Baruch; Gonzalez, Roxana M.; Lerner, Jennifer S.; Small, Deborah A. (2005) "Evolving Judgments of Terror Risks: Foresight, Hindsight, and Emotion". *Journal of Experimental Psychology: Applied*, 11(2), 124-139.

Flynn, J., Slovic, P., and C.K. Mertz (1994), "Gender, Race, and Perception of Environmental Health Risk", *Risk Analysis*, 14 (6), 1101-1108.

Gustafson, P.E. (1998), "Gender Differences in Risk Perception: Theoretical and Methodological Perspectives", *Risk Analysis*, 18(6), 805-811.

Helweg-Larsen, M, (1999), "The Lack of Optimism Biases in Response to the 1994 Northridge Earthquake: the Role of Personal Experience", *Basic Applied Social Psychology*, 21, 119-129.

Juric, Pavica (2006), "Mass Media Usage during a Natural Disaster: LSU College Students and Hurricane Katrina", unpublished Masters thesis, School of Mass Communication, Louisiana State University.

Knobb, Richard D., Rhome, Jamie R., and Richard D. Brown (2005) "Tropical Cyclone Report: Hurricane Katrina, 23-30 August 2005", Technical Report, NOAA National Hurricane Center.

Ma, H. and Carlin, B.P., (2007) "Bayesian Multivariate Areal Wombling for Multiple Disease Boundary Analysis," *Bayesian Analysis*, 2, 281-302.

NOAA National Hurricane Center (2005), Advisory Archive
<http://www.nhc.noaa.gov/archive/2005/index.shtml>

Norris, F. H., Smith, T., and K. Kaniasty (1999), "Revisiting the Experience-Behavior Hypothesis: the effects of Hurricane Hugo on Hazard Preparedness and other Self-Protective Acts", *Basic Applied Psychology*, 21, 27-47

Peacock, Walter G., Brody, Samuel D., and Wes Highfield (2005), "Hurricane Risk Perceptions Among Florida's Single-Family Homeowners", *Landscape and Urban Planning*, 73, 120-135.

Peacock, Walter G. (2003), "Hurricane Mitigation Status and Factors Influencing Mitigation Decisions Among Florida's Single-Family Homeowners", *Natural Hazards Review*, 4(3), 1-10.

Perez-Lugo, Marla (2004), "Media Uses in Disaster Situations: A New Focus on the Impact Phase", *Sociological Inquiry*, 74(2), 210-225.

Perry, R.W., Lindell, M.K., and Greene, M.R. (1982), "Threat Perceptions in Public Response to Volcano Hazard", *Journal of Social Psychology*, 116, 199-204.

Pew Internet Life Project (2008);

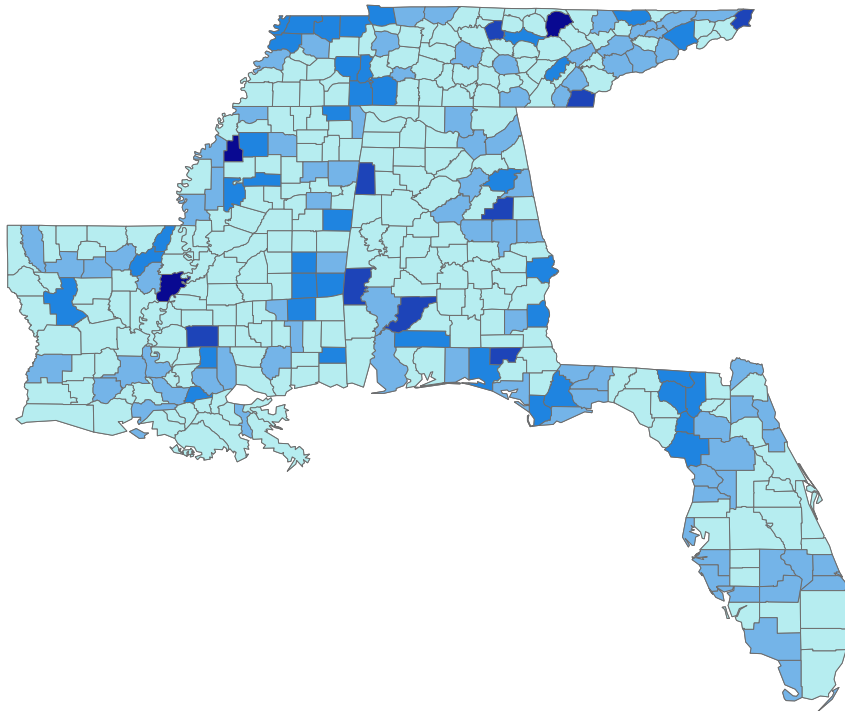
http://www.pewinternet.org/trends/User_Demo_10%2020%2008.htm.

Piotrowski, C. and T. Armstrong (1998), "Mass Media Preferences in Disaster: A Study of Hurricane Danny", *Social Behavior and Personality*, 26(4), 341-6.

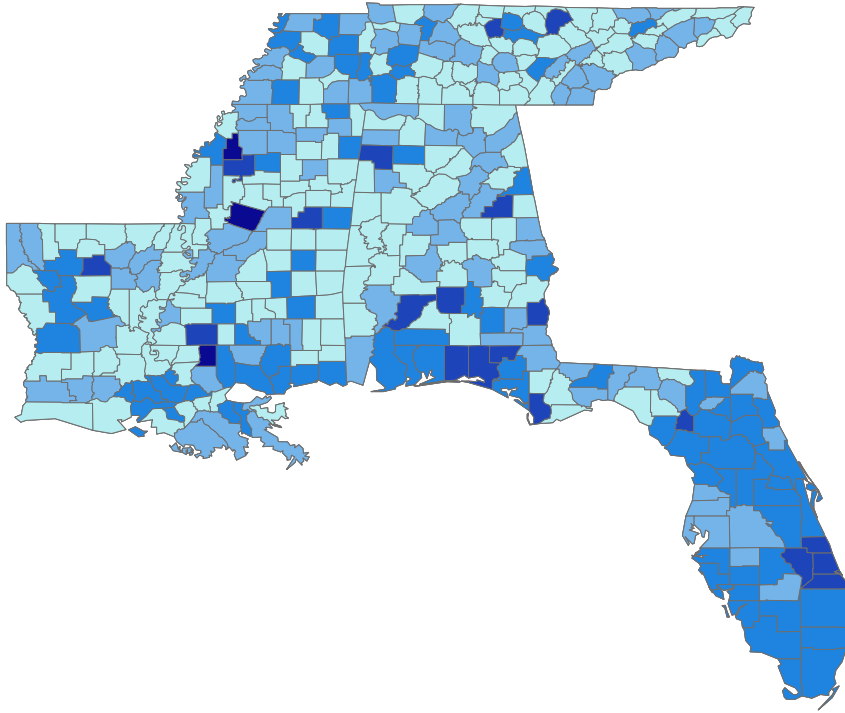
Ramsay, J.O. and Silverman, B.W. (2005), *Functional Data Analysis*. Springer.

Figure 1: Mean Hourly Relative Frequency of Visitation Rates to Weather Sites by County

1a: August 19th, four days before Katrina. A darker color indicates higher frequencies of traffic at weather-related sites (see legend below). Note high baseline visitation rates in hurricane-prone coastal areas of Florida and the Gulf and inland urban areas.



1b: August 24th, Katrina forms east of Florida and Warnings are issued for South Florida. Note increase in visitation throughout the entire Florida Peninsula



1c: August 27th: Katrina is in the Southeast Gulf and Warnings are issue for the northern Gulf Coast. Note increased traffic levels along the Gulf coast, and sustained high traffic throughout the Florida peninsula.

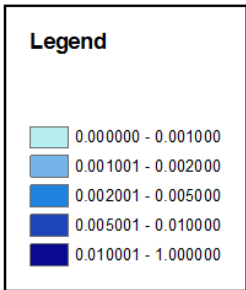
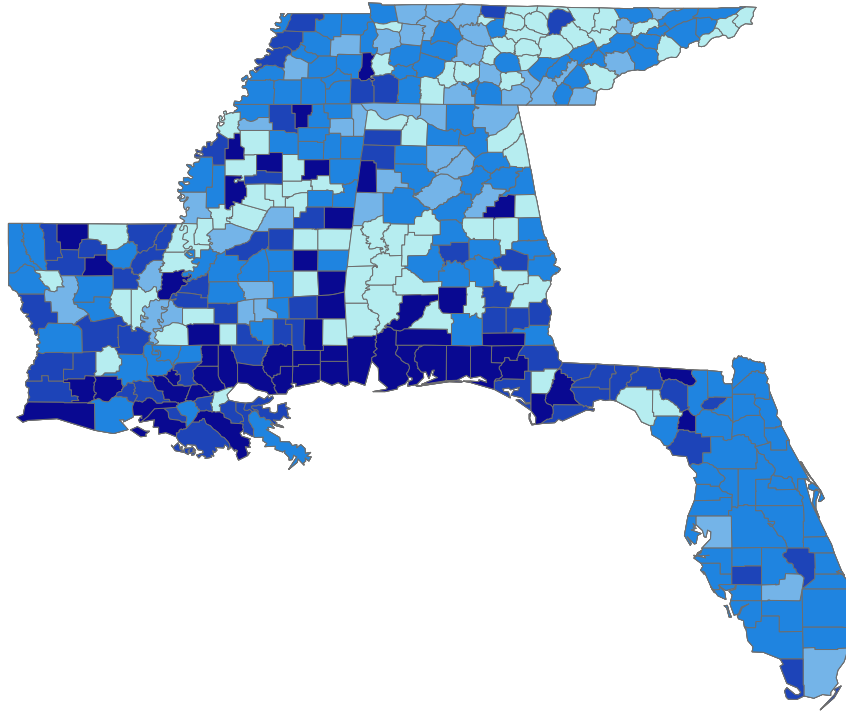
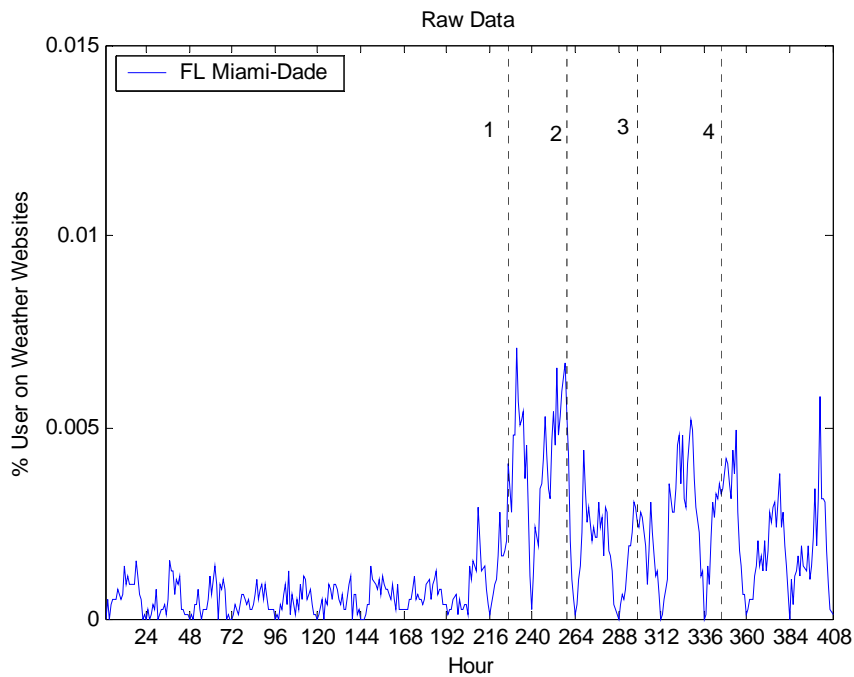
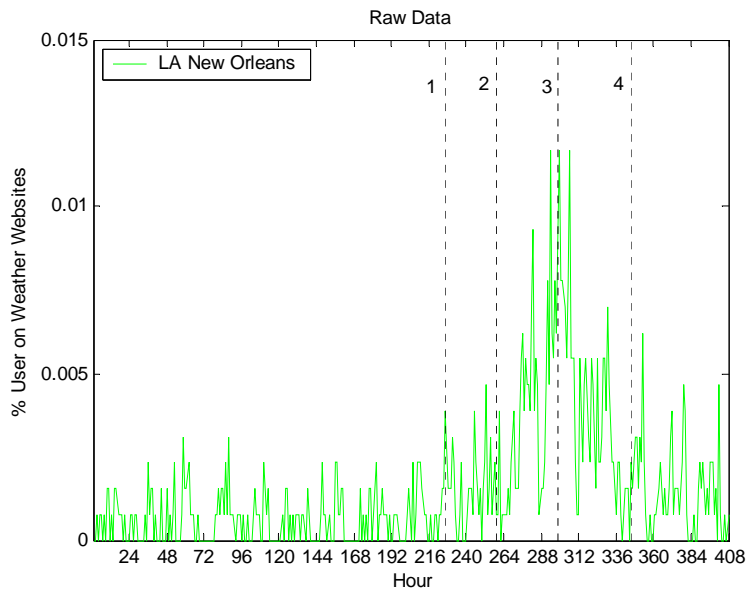


Figure 2: Hourly Weather Site Visitation Rates for Miami and New Orleans

2a: Miami-Dade, FL*



2b: New Orleans, LA*



* The first event (vertical dashed line) indicates the time of the hurricane watch for southeast Florida. Line 2 corresponds to when Katrina made first landfall on Florida. Line three corresponds to when a hurricane watch for the southeastern coast of Louisiana, including New Orleans. The fourth indicates when Katrina passed just to the east of New Orleans.

2c Graphical Display of Storm Positions Relative to the Time Events in 2a and 2b*

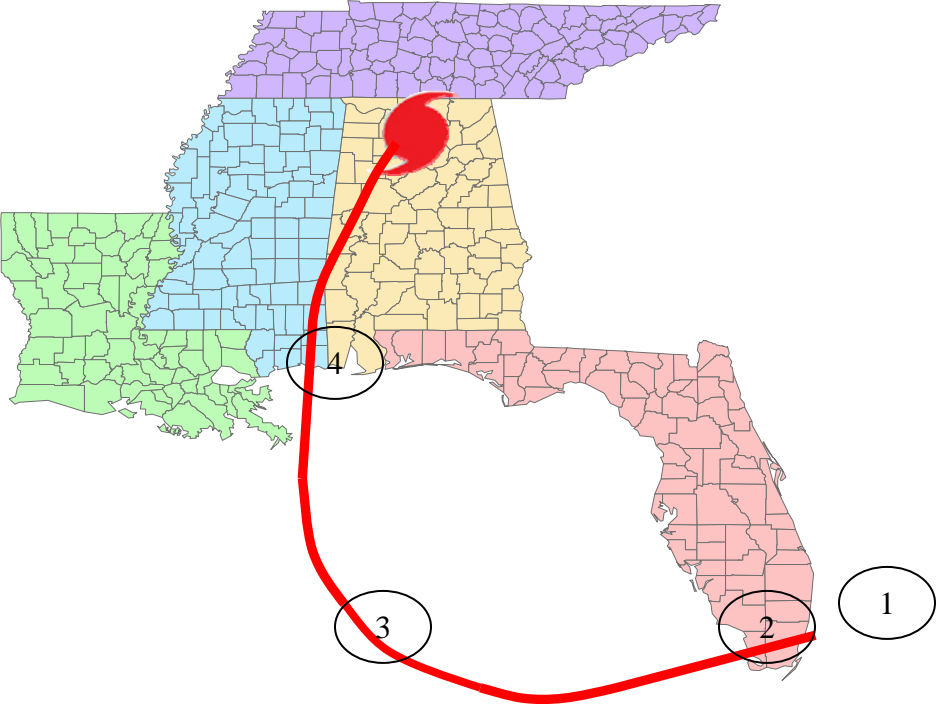


Figure 3: Hurricane Risk Zones Used for the Analysis

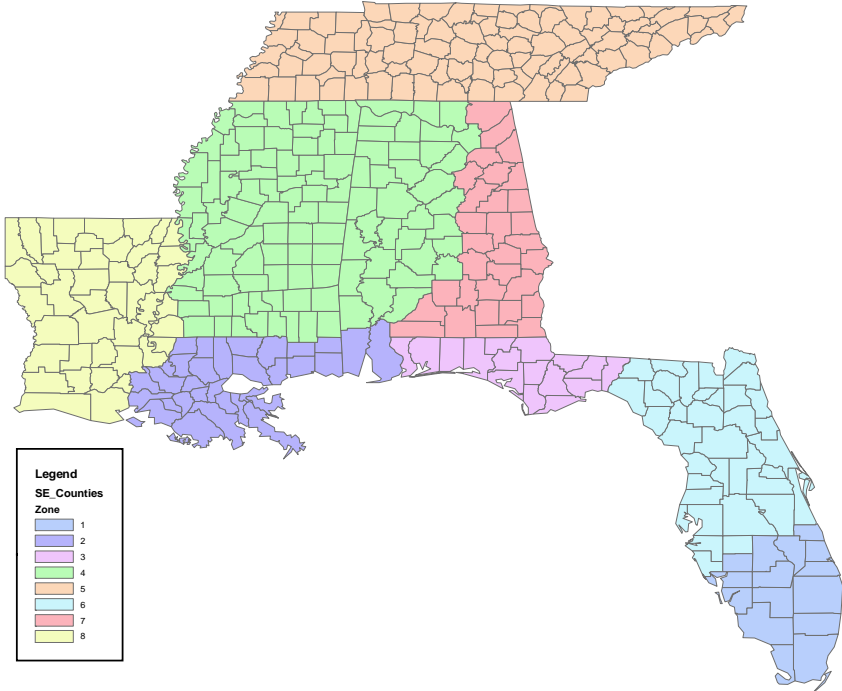
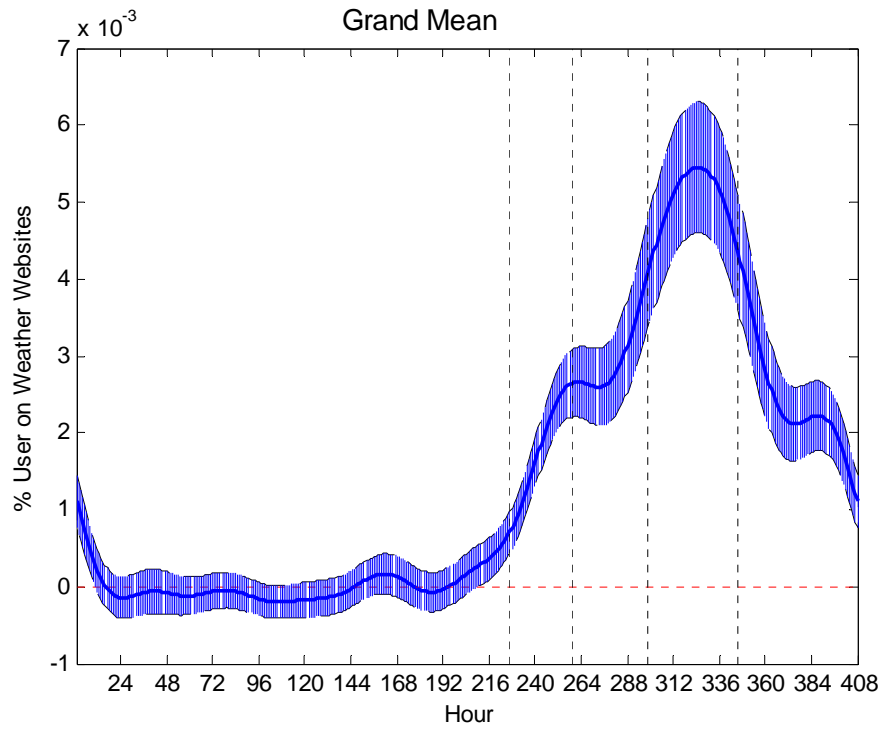
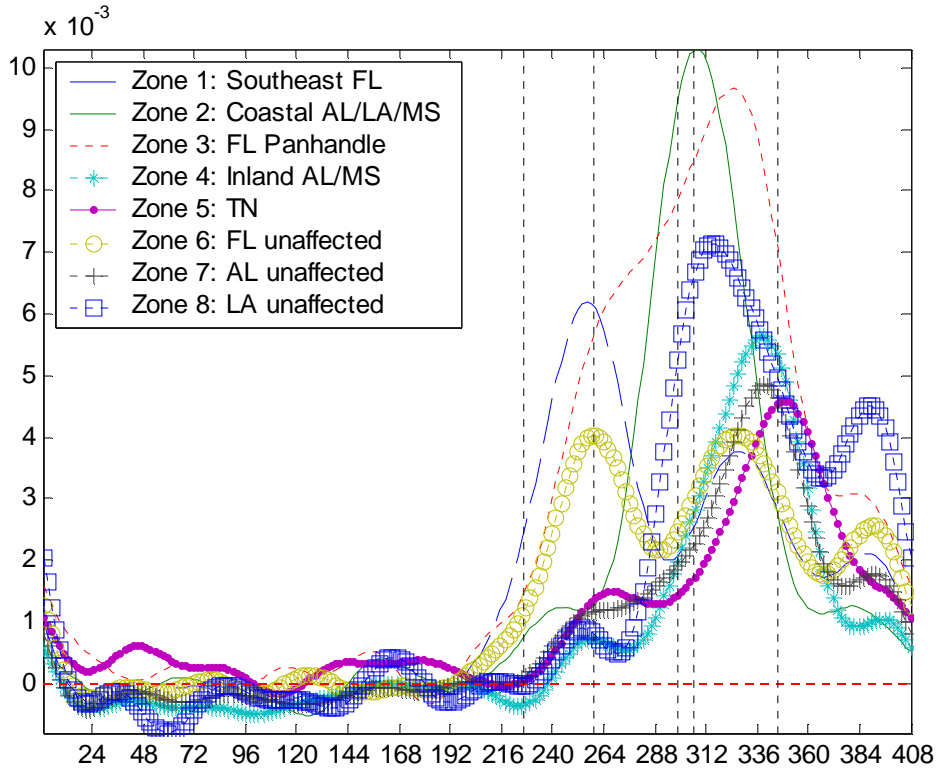


Figure 4: Grand Mean Function with 95% Confidence Interval



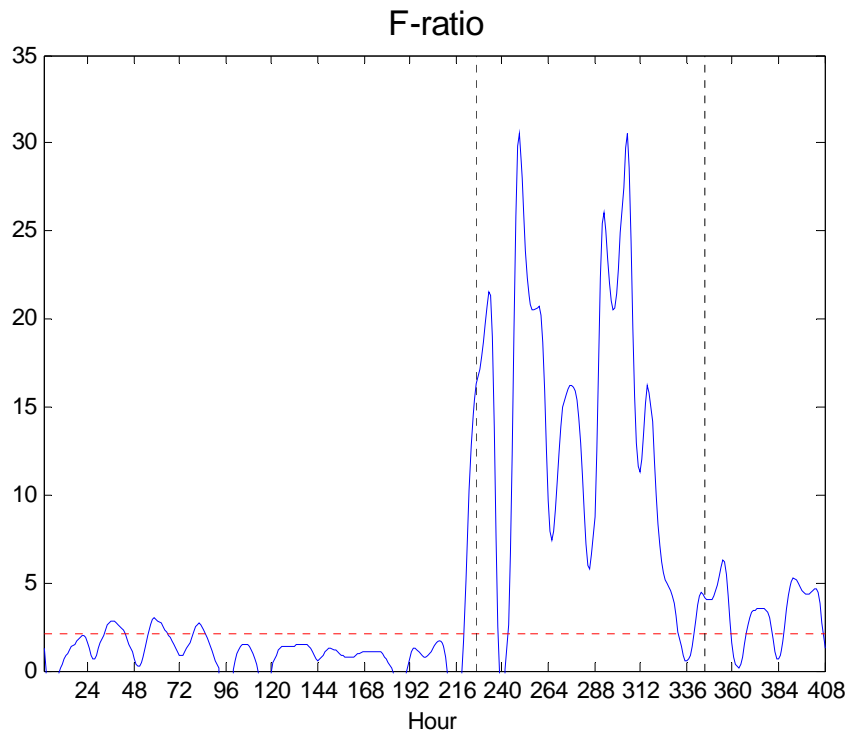
* The first vertical dashed line indicates the time of the hurricane watch for southeast Florida. The second vertical dashed line indicates the time Katrina made first landfall on Florida. The third vertical dashed line indicates the time of the hurricane watch for the southeastern coast of Louisiana, including New Orleans. The fourth vertical dashed line indicates the time that Katrina passed by New Orleans.

Figure 5: Zone Specific Effect Functions



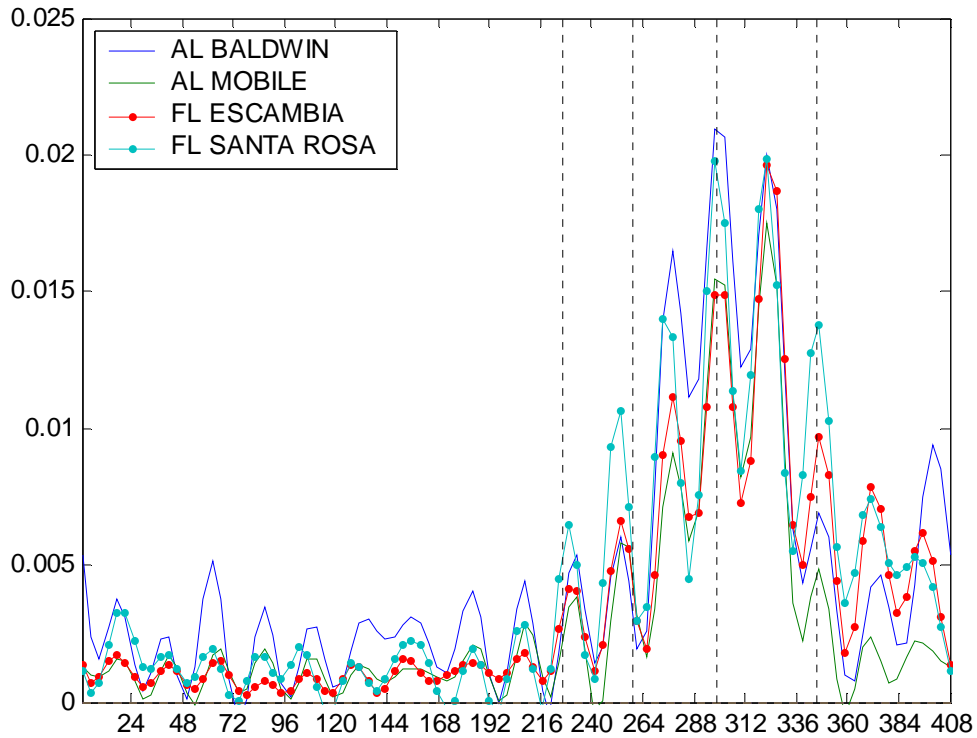
* The first vertical dashed line indicates the time of the hurricane watch for southeast Florida. The second vertical dashed line indicates the time Katrina made first landfall on Florida. The third vertical dashed line indicates the time of the hurricane watch for the southeastern coast of Louisiana, including New Orleans. The fourth vertical dashed line indicates the time that New Orleans mayor declared a state of emergency and issued a voluntary evacuation order. The fifth vertical dashed line indicates the time that Katrina passed by New Orleans.

Figure 6: F-ratio Functions



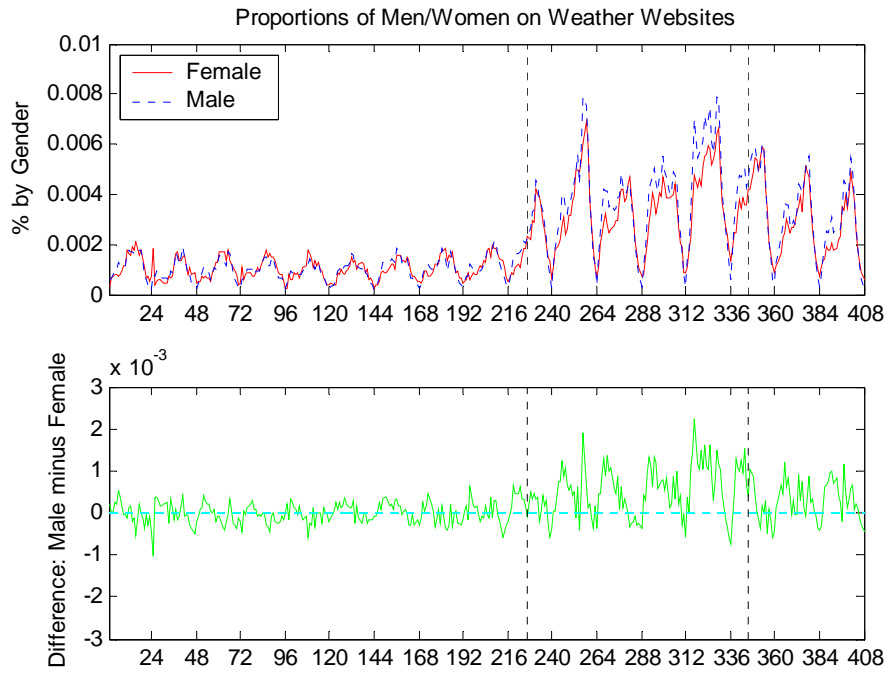
* The first vertical dashed line indicates the time of the hurricane watch for southeast Florida. The second vertical dashed line indicates the time that Katrina passed by New Orleans.

Figure 7: Weather Website Visitation Patterns for Four Contiguous Gulf-Coast Counties in Florida and Alabama

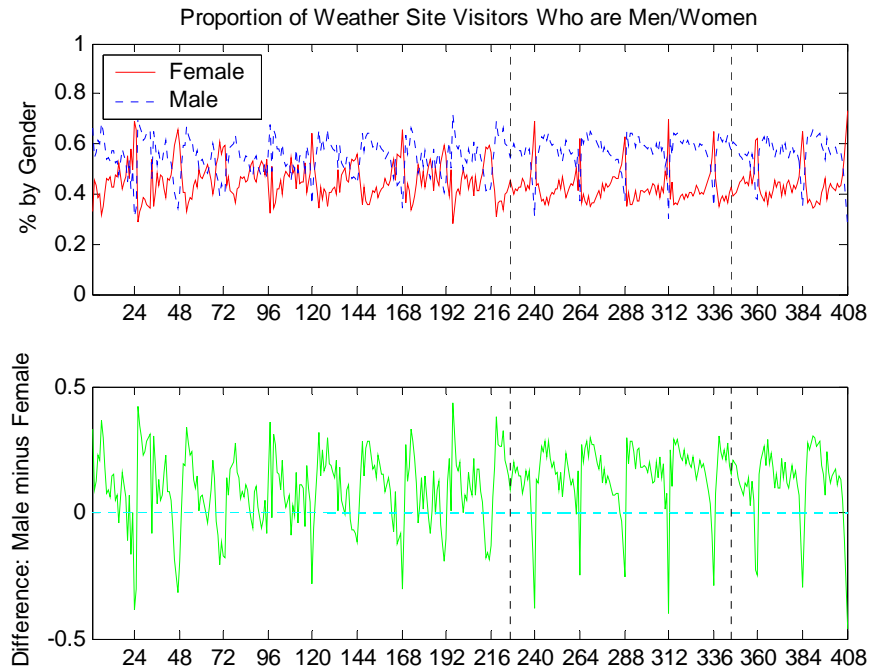


* The first vertical dashed line indicates the time of the hurricane watch for southeast Florida. The second vertical dashed line indicates the time Katrina made first landfall on Florida. The third vertical dashed line indicates the time of the hurricane watch for the southeastern coast of Louisiana, including New Orleans. The fourth vertical dashed line indicates the time that Katrina passed by New Orleans.

Figure 8: Gender Effects on Weather Website Visitation
8a Visitation Probability Conditional on Gender

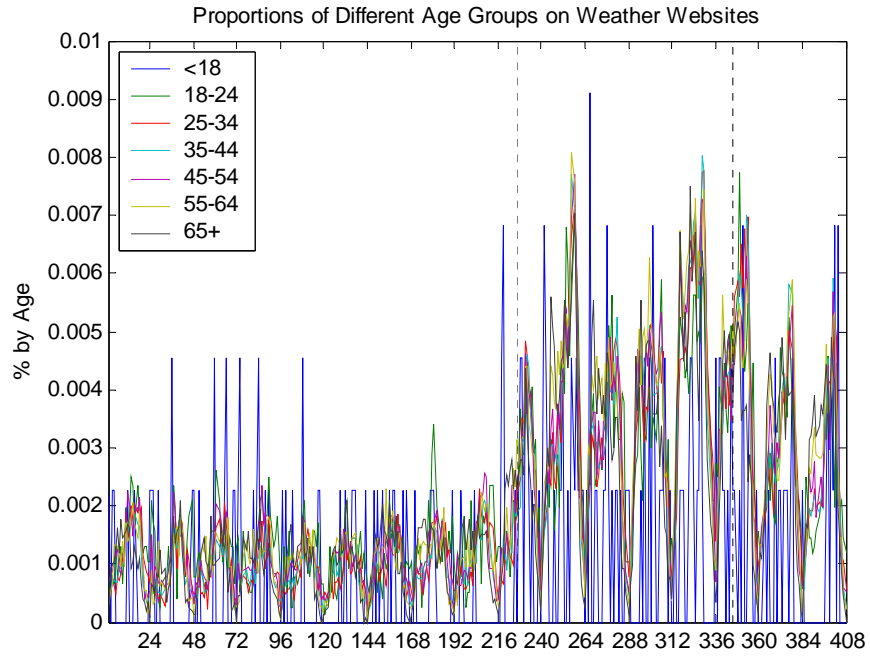


8b Gender Conditional on Visitation



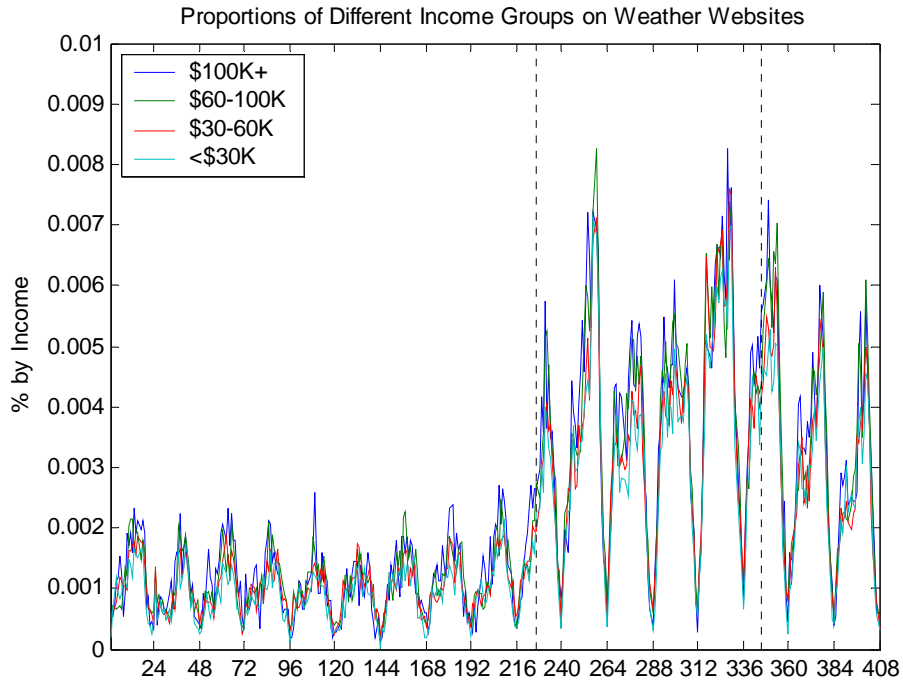
* The first vertical dashed line indicates the time of the hurricane watch for southeast Florida. The second vertical dashed line indicates the time that Katrina passed by New Orleans.

Figure 9: Age Effects on Weather Website Visitation - Visitation Probability Conditional on Age



* The first vertical dashed line indicates the time of the hurricane watch for southeast Florida. The second vertical dashed line indicates the time that Katrina passed by New Orleans.

Figure 10: Income Effects on Weather Website Visitation - Visitation Probability Conditional on Income



* The first vertical dashed line indicates the time of the hurricane watch for southeast Florida. The second vertical dashed line indicates the time that Katrina passed by New Orleans.

Figure 11: Overall Visitation Traffic to FEMA-type Sites by Day

